

Anexo Metodológico

Introducción

Uno de los principales objetivos del *Informe Estado de la Nación* es proveer al país información oportuna, que permita conocer el avance en el logro de sus aspiraciones de desarrollo humano sostenible. Con la intención de aumentar la profundidad con que se analizan los temas, cada año interviene una amplia red de instituciones e investigadores, quienes colaboran con el suministro de datos actualizados y la aplicación de técnicas y mediciones novedosas, que facilitan una comprensión más objetiva de la realidad nacional. Al incorporar distintos instrumentos metodológicos se pretende dar una sólida base técnica a los hallazgos presentados en cada capítulo.

En este “Anexo Metodológico” se exponen los procedimientos técnicos seguidos para el abordaje de algunos temas incluidos en esta edición. En el capítulo “Equidad e integración social” fueron los siguientes: i) patrones territoriales y factores sociodemográficos asociados a los homicidios y la incautación de drogas, ii) factores que inciden en la inserción laboral de mujeres y hombres, iii) modelo predictivo de movilidad laboral, y iv) discriminación salarial por motivos de género.

Tres temas corresponden al capítulo “Oportunidades, estabilidad y solvencia económicas”, a saber: i) estimación del ajuste fiscal requerido para estabilizar el crecimiento de la deuda pública, ii) proyección y descomposición de la tasa de desempleo en Costa Rica, y iii) determinantes de la estructura productiva que impactan la desigualdad.

En el capítulo especial, titulado “Transporte y movilidad: retos en favor del desarrollo humano” se efectuaron los siguientes procesos: i) estimación de los tiempos de viaje de las personas trabajadoras en la Gran Área Metropolitana (GAM), ii) estimación de los costos derivados del congestionamiento vial para las personas que se movilizan de un cantón a otro de la GAM para trabajar y iii) análisis espacial sobre la congestión vehicular a partir de la base de datos de la aplicación Waze.

Aportes metodológicos en materia de equidad e integración social

Patrones territoriales y factores sociodemográficos asociados a los homicidios y la incautación de drogas en Costa Rica

En el marco de un convenio de colaboración entre el Instituto Costarricense sobre Drogas (ICD) y el Programa Estado de la Nación, Sánchez (2018a) analizó la presencia de patrones territoriales en los homicidios y la incautación de drogas en el país. Para ello, utilizó una metodología basada en sistemas de información geográfica (SIG), la cual consiste en la integración de datos geográficamente referenciados y su procesamiento a través de técnicas de Estadística Espacial.

El objetivo del estudio fue identificar la existencia de aglomeraciones de las unidades geográficas de observación (distritos) en torno a los valores exhibidos por un atributo de interés (homicidios e incauciones de drogas), entendidos

como variables que describen características específicas de esas unidades.

Para determinar si existe dependencia espacial, es decir, si los valores de un indicador en un distrito son parcialmente función del valor de ese mismo indicador en distritos vecinos, se usó la técnica de análisis de *clusters* y de valor atípico, denominada *I Anselin local de Moran*. La dependencia espacial significa que los valores de una misma variable (homicidios o incautación de drogas), que son medidos en localidades (distritos) cercanas entre sí, tienden a ser similares.

Esta herramienta detecta *clusters* o aglomeraciones espaciales de las variables analizadas que presentan valores altos o bajos, así como los valores atípicos. Primero se elabora el SIG y luego se calcula un valor *I de Moran local*, una puntuación *z*, un pseudovalor *P* y un código que representa el tipo de *cluster* para cada entidad estadísticamente significativa. Las puntuaciones *z* y los pseudovalores *P* representan la significancia estadística de los valores estimados.

Con el cálculo de la *I de Moran local* se obtienen los siguientes resultados: i) un valor positivo para *I* indica que ese distrito tiene vecinos con valores altos o bajos similares; entonces, este distrito es parte de un *cluster*; ii) un valor negativo para *I* indica que el distrito tiene vecinos con valores distintos, por lo que es una entidad con un valor atípico. En ambos casos el valor *P* debe ser lo suficientemente pequeño para que se considere significativo desde el punto de vista estadístico.

La combinación descrita configura cuatro grupos:

- un *cluster* estadísticamente significativo de valores altos (AA),
- un *cluster* de valores bajos (BB),
- un valor atípico, en el que un valor alto está rodeado principalmente por valores bajos (AB) y,
- un valor atípico, en el que un valor bajo está rodeado principalmente por valores altos (BA).

La importancia estadística se estableció en un nivel de confianza del 95%. Los distritos con valores P inferiores a 0,01 se consideran estadísticamente significativos. La información de base para este ejercicio fue suministrada por el ICD, con datos de la Policía de Control de Drogas (PCD).

Los resultados confirmaron la existencia de autocorrelación espacial, tanto en la tasa de homicidios como en las incautaciones de drogas, es decir, estas variables no se comportan de manera aleatoria, sino que presentan patrones de aglomeración territorial.

Para explorar las variaciones espaciales de algunos factores que se podrían asociar a los homicidios y la incautación de drogas, Sánchez (2018a) estimó regresiones ponderadas geográficamente, denominadas GWR por su sigla en inglés (Fotheringham et al., 2002). El uso de GWR es una técnica que proporciona un modelo local de la variable que intenta entender/prever al ajustar una ecuación de regresión a cada entidad en el *dataset*.

Dado que se tenía información desagregada a nivel distrital, y que ambas problemáticas varían en el territorio y dependen de múltiples factores, se optó por desarrollar modelos globales (para todo el país) y locales (por distritos).

Los parámetros estimados por los modelos globales se interpretan como un valor promedio a nivel nacional, es decir, se asume que un factor determinado afecta de manera similar a todos los distritos, lo cual resulta en un supuesto fuerte dadas las diferencias territoriales.

Con las GWR se construye una ecuación

distinta para cada entidad del *dataset* (distrito), mediante la incorporación de las variables dependientes y explicativas para cada caso. El modelo se corre en el programa ArcGIS, que es un conjunto de productos de *software* utilizado en el campo de los SIG.

Las variables dependientes empleadas en los modelos fueron: i) la tasa de homicidios por cada 10.000 habitantes en el período 2014-2016 y ii) la cantidad absoluta de incautaciones de drogas, entre 2015 y 2017. La unidad geográfica de análisis, como ya se mencionó, fue el distrito.

Las variables independientes se clasificaron en dos dimensiones: i) socioeconómicas, que incluyen: porcentaje de hogares con necesidades básicas insatisfechas (NBI), tasa de desempleo abierto, porcentaje de población que vive en asentamientos informales (precarios), porcentaje de población migrante nicaragüense, años promedio de escolaridad y un índice de concentración económica en turismo¹; ii) demográficas: edad promedio de los residentes del distrito, porcentaje de población masculina, densidad (habitantes por kilómetro cuadrado), porcentaje de población urbana, porcentaje de población joven (15 a 30 años), tasa de dependencia infantil, porcentaje de madres adolescentes solteras y porcentaje de población de entre 18 y 30 años que no ha concluido el tercer año de la educación secundaria. No se utilizó la desigualdad de ingresos (medida con el coeficiente de Gini) debido a que no está disponible con este nivel de desagregación.

Los indicadores se calcularon con el Censo Nacional de Población de 2011, del INEC. Algunas variables demográficas provienen del Centro Centroamericano de Población de la Universidad de Costa Rica (CCP-UCR).

Una de las principales limitaciones del estudio es que no se tienen datos a nivel distrital relacionados con los factores disuasorios del homicidio y el narcotráfico. Según la teoría de la economía del delito, las personas responden a incentivos y castigos, así como a los costos y beneficios de delinquir, por lo que aspectos como la presencia y efectividad de la policía, la dureza de las penas y la

probabilidad de ir a la cárcel influyen en la decisión de transgredir la ley. Tampoco se cuenta con información sobre factores asociados a las políticas públicas sobre seguridad ciudadana (Sánchez, 2018a).

Para determinar la existencia de correlación espacial en los modelos planteados se siguieron dos pasos. Primero se corrió un modelo con las variables. Luego se calculó nuevamente el estadístico de Moran. Al estimar el modelo inicial se descartó una serie de variables que presentaban bajos niveles de significancia, heterocedasticidad² y multicolinealidad³. Después se aplicaron las pruebas de multiplicadores de Lagrange para definir cuál modelo (de rezago o de error) era mejor para la estimación.

Los resultados obtenidos dieron un estadístico de Moran positivo y significativo para ambos modelos. Las pruebas del multiplicador de Lagrange, tanto para el modelo de error como para el de rezago, fueron significativas. De acuerdo con la prueba del multiplicador robusto, el mejor modelo para estimar los factores determinantes es el de rezago espacial.

Los coeficientes de las variables explicativas y el nivel de ajuste de cada uno de los modelos se puede consultar en el estudio de Sánchez (2018a), disponible en el sitio www.estadonacion.or.cr



PARA MÁS INFORMACIÓN SOBRE
**PATRONES TERRITORIALES EN
HOMICIDIOS Y DROGAS**

véase Sánchez, 2018a, en
www.estadonacion.or.cr

**Modelo para determinar factores
que inciden en la inserción laboral de
mujeres y hombres**

Para conocer los factores demográficos y socioeconómicos que inciden en que una persona participe o no en la fuerza de trabajo, Morales y Segura (2018) estimaron con el modelo de regresión logística⁴ que se muestra en la ecuación 7.1.

El cálculo se realizó para hombres y mujeres entre 25 y 60 años de edad, durante el período 2001-2017, con el fin de analizar la evolución de los factores

ECUACIÓN 7.1

Modelo para determinar factores que inciden en la inserción laboral de mujeres y hombres

$$P(PL) = \beta_0 + \beta_1(\text{edad}) + \beta_2(\text{escolaridad}) + \beta_3(\text{J.monoparental}) + \beta_4(\text{J.pareja/hijos}) + \beta_5(\text{J.pareja/nohijos}) + \beta_6(\text{J.calificada}) + \beta_7(\text{niños}) + \beta_8(\text{Z.urbana}) + \beta_9(\text{R.Central}) + \beta_{10}(\text{educ.noregular}) + \beta_{11}(\text{otra.mujer})$$

Donde:

P (PL): probabilidad de participación laboral (1 participa, 0 no participa).

Edad: variable continua de la edad en años.,

Escolaridad: variable continua de la cantidad de años de educación formal que aprobó.

J.monoparental: variable *dummy* para jefa o jefe de hogar monoparental, es decir, jefaturas sin pareja, pero con hijos (1: jefe monoparental, 0: otro caso).

J.pareja/hijos: variable *dummy* para jefa o jefe de hogar que tiene pareja e hijos (1: jefe con pareja e hijos, 0: otro caso).

J.pareja/nohijos: variable *dummy* para jefa o jefe de hogar que tiene pareja pero no hijos (1: jefe con pareja sin hijos, 0: otro caso).

J.calificada: variable *dummy* para jefa o jefe de hogar que tiene la secundaria completa o más como logro educativo (1: jefe calificado, 0: otro caso).

Niños: variable continua con el número de niños menores de 7 años en el hogar.

Z.Urbana: variable *dummy* de residencia en zona urbana (1: urbana, 0: rural).

R.Central: variable *dummy* de residencia en la región Central (1: Central, 0: resto).

Educ.noregular: variable *dummy* del caso en que la persona cursó educación no regular (1: sí, 0: no).

Otra.mujer: variable *dummy* del hogar en que reside más de una mujer de 18 años y más (1: sí, 0: no).

en el tiempo. Se usaron las encuestas de hogares del INEC. Los resultados de las regresiones (coeficientes, *odds ratio* y errores estándar) están disponibles en la ponencia de Morales y Segura (2018), en el sitio www.estadonacion.or.cr

Modelo predictivo de movilidad laboral por género

El capítulo “Equidad e integración social” documentó que Costa Rica tiene un margen para ampliar la participación femenina en el mercado laboral. En este sentido, es relevante conocer en cuáles sectores económicos podrían trabajar las mujeres, qué oportunidades tendrían para movilizarse hacia otras actividades mejor remuneradas y cuál sería el impacto de esa inserción sobre la incidencia de la pobreza total en el país.

Para aproximar estos temas, Morales y

Segura (2018) utilizaron una metodología que se basa en métodos predictivos depurados con técnicas de aprendizaje supervisado (*machine learning*). Consiste en realizar múltiples ejecuciones del mismo algoritmo, o bien contrastar diferentes modelos con variantes en sus parámetros, a fin de obtener el mejor modelo posible para los datos existentes. También se somete ese modelo a “pruebas de estrés”, dado que de manera recurrente se pueden omitir tramos de la base de datos, para verificar si es poco robusto ante variaciones en la información. El modelo seleccionado debe tener tasas de error muy bajas ante cualquier conjunto de datos.

Para analizar la movilidad laboral se definieron cinco sectores económicos: agropecuario, industria tradicional, nueva industria (empresas de exportación en zonas francas), servicios y sector público. Y para conocer el perfil que determina la pertenencia a cada sector se consideraron dieciséis variables disponibles en la serie histórica de las encuestas de hogares del INEC para el período 2001-2017, las cuales se especifican a continuación:

- Capacidades: combina el nivel educativo y el dominio de un segundo idioma.
- Sector: segmento económico en el cual trabaja la persona, privado o público.
- Edad: en grupos decenales.
- Escolaridad: años acumulados por cada nivel de educación aprobado.
- Lugar de residencia: establece si la persona vive en la región Central o en resto del país, por zona urbana o rural.
- Educación no regular: indica si la persona recibió o no educación no formal, es decir, cursos que no forman parte del ciclo educativo regular.
- Aseguramiento: indica si la persona está o no cubierta por el seguro de la CCSS.
- Clase social: variable creada para determinar el estrato social de una persona,



PARA MÁS INFORMACIÓN SOBRE
MODELOS DE INSERCIÓN LABORAL
véase Morales y Segura, 2018
www.estadonacion.or.cr

según su rol en el mercado laboral, oficio y educación (para más detalles de sus categorías, véase el *Decimoquinto Informe Estado de la Nación*).

- **Tamaño del hogar:** cantidad de personas que conforman el hogar.
- **Estructuras internas de los sectores económicos:** serie de cinco variables que resumen el porcentaje de personas ocupadas miembros del hogar que trabajan en el sector indicado.
- **Responsabilidades:** condiciones domésticas que le restan a una mujer las posibilidades de incursionar en el mercado o movilizarse a otros sectores. Indica si las jefas, cónyuges mujeres e hijas tienen hijos, nietos o adultos mayores que requieren cuidado.

La población se delimitó para el rango de edad de 25 a 65 años, con el fin de descartar estudiantes, menores de edad y personas pensionadas.

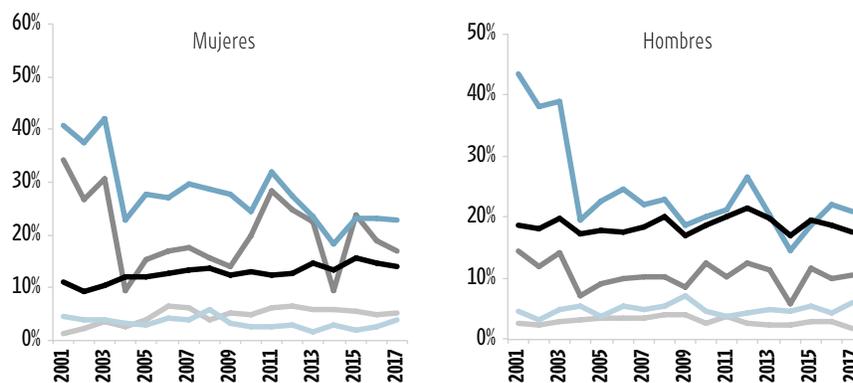
Para evaluar la robustez de los modelos se realizaron y compararon las predicciones utilizando seis métodos: bayesiano, máquinas de soporte vectorial, árboles de decisión, bosques aleatorios, modelos de potenciación *AdaBosting* y la predicción según el vecino más cercano. Los resultados indicaron que las predicciones bayesianas son las más precisas para este conjunto de datos. La tasa global de clasificación correcta es de 86%, distribuida de la siguiente manera: 96% en el sector agropecuario, 79% en la industria tradicional, 72% en la nueva industria, 88% en los servicios y 96% en el sector público. Estas cifras muestran un ajuste más que satisfactorio.

El modelo predictivo también permite calcular cuántas personas ocupadas, por su perfil, tienen posibilidades de trasladarse a un sector productivo distinto a aquel en que trabajan. Para ello se efectuaron simulaciones suponiendo ciertas condiciones que pueden potenciar la movilidad, en concreto las siguientes: i) las personas se hacen bilingües, es decir, se les asigna el dominio de un segundo idioma; ii) se asume que las personas que no completaron la educación primaria o secundaria sí concluyeron el

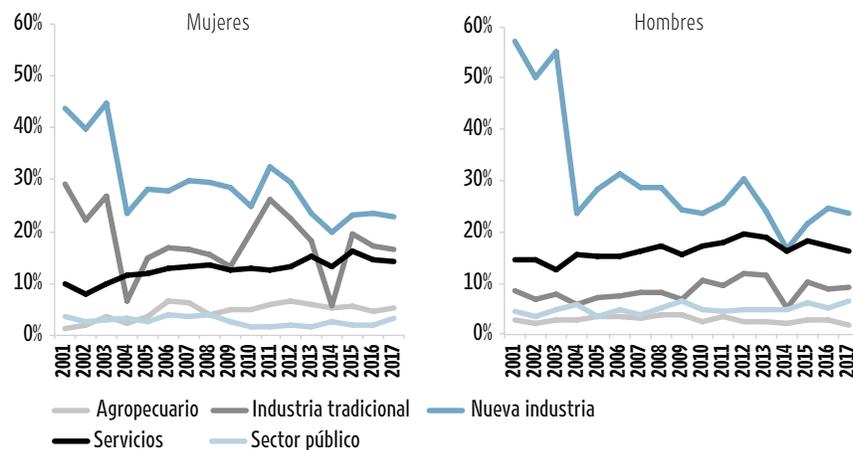
GRÁFICOS 7.1

Porcentaje de personas ocupadas que se pueden movilizar hacia otros sectores, según escenarios simulados^{a/}

a) Primera simulación: la persona se hace bilingüe



b) Segunda simulación: las personas con niveles incompletos terminan la primaria o la secundaria



a/ La simulación consiste en predecir el sector económico más factible, en función de las características de las personas y el perfil que demanda cada actividad. Luego se confronta la predicción con el sector en que realmente están trabajando. Cuanto mayor es el porcentaje de movilización en un sector, mayor es la cantidad de personas cuyos perfiles que se ajustan a la demanda de otras actividades.

Fuente: Morales y Segura, 2018, con datos de las encuestas de hogares, del INEC.

nivel respectivo; iii) se supone que las mujeres que trabajan no tienen personas dependientes. Se hicieron combinaciones entre estas variables, para un total de siete escenarios. Para confirmar la movilidad, se confrontó la predicción que estimó el modelo para cada persona, con el sector real declarado en la encuesta. Los gráficos 7.1 muestran los resultados de dos escenarios simulados, pues no hay muchas diferencias en los resultados obtenidos en los siete casos considerados.

El modelo predictivo de los ocupados

sirve para pronosticar los sectores en los que tendrían más probabilidades de insertarse las personas desocupadas o que están fuera de la fuerza de trabajo, según sus características sociodemográficas. En el modelo se excluyeron dos variables del perfil: el sector institucional de pertenencia (público o privado) y la clase social, pues los datos solo están disponibles para la población ocupada. Nuevamente se evaluaron los modelos con técnicas de aprendizaje y validaciones cruzadas. En este caso se determinó que

el mejor método de pronóstico es la técnica denominada “bosques aleatorios”. Los datos obtenidos estiman que este algoritmo tiene un ajuste correcto global del 97%, distribuido del siguiente modo: 95% en el sector agropecuario, 87% en la industria tradicional, 96% en la nueva industria, 98% en servicios y 99% en el sector público.

Una vez realizadas las predicciones, se asignó a las personas que actualmente están fuera del mercado laboral o buscan empleo y no lo consiguen, el ingreso promedio que perciben las personas que ya están insertas en los sectores respectivos. Con estos ingresos se recalculó la condición de pobreza en los hogares y se midió el impacto que tendría el empleo en la pobreza nacional. Los resultados se pueden consultar en la ponencia de Morales y Segura (2018), disponible en el sitio www.estadonacion.or.cr.



PARA MÁS INFORMACIÓN SOBRE **INSERCIÓN Y MOVILIDAD LABORAL** véase Morales y Segura, 2018, en www.estadonacion.or.cr

Discriminación salarial por motivo de género en Costa Rica

La técnica Oaxaca-Blinder es un método que permite analizar diferencias en los ingresos, popularizado por los economistas Oaxaca (1973) y Blinder (1973). Solórzano (2018) lo usó para estimar el monto que representa la discriminación salarial en la brecha de ingresos entre sexos.

Esta técnica subdivide los ingresos promedio de dos grupos demográficos. El primer componente muestra las diferencias en el nivel de calificación, las cuales son explicadas por el modelo. El segundo evidencia las brechas en la estructura del modelo, es decir, estima la discriminación en el mercado de trabajo. En Economía, el método de Oaxaca-Blinder ha sido ampliamente utilizado para realizar descomposiciones salariales y determinar cuánto de la diferencia se debe a discriminación y cuánto a disparidades en la dotación de capital humano.

Según Fuentes et al. (2005), es necesario modelar los salarios en función del capital humano, para luego poder descomponer la brecha salarial en dos elementos: uno representado por las diferencias en las características de ambos grupos y otro que muestra la disparidad generada por el mercado. En un mercado laboral libre de discriminación, el segundo elemento debería ser idéntico para hombres y mujeres, y las desigualdades salariales se deberían únicamente a diferencias en la dotación de capital humano. Si existiese discriminación, el retorno sería distinto entre los grupos, por lo que la diferencia sería el factor discriminatorio. Por ejemplo, existiría inequidad salarial si por cada año de escolaridad adicional los hombres reciben una retribución mayor que las mujeres, por el hecho de ser hombres.

Al estimar los factores que inciden en los salarios se incurre en un sesgo de selección, por dos razones. Por un lado, la muestra utilizada está compuesta solo por las personas que tenían trabajo al momento de la encuesta y, por ende, se conoce su salario; por otro lado, existe segregación ocupacional, es decir, discrepancias asociadas a la elección de una ocupación determinada. Este sesgo afecta las diferencias salariales, debido a que las ocupaciones tienen salarios promedio distintos y existen barreras de entrada a ciertos empleos para algunos grupos (las mujeres).

A continuación se desarrolla en notación matemática la técnica aplicada. Si se denotan dos grupos como A y B, la diferencia entre resultados de medias ($\Delta\bar{Y}$) se resume como la diferencia entre la media de las observaciones del grupo A y del

B. La descomposición de tres términos se presenta en la ecuación 7.2 (Oaxaca-Blinder).

La descomposición del método Oaxaca-Blinder divide las diferencias de medias en una porción que obedece a brechas entre grupos en las variables explicativas y una parte que no se puede explicar. Esta última es comúnmente atribuida a discriminación, pero podría ser el resultado de variables no observadas en el modelo. Además, el componente no explicado se puede dividir en dos partes: la discriminación a favor del grupo A y la discriminación en contra del grupo B.

Para las estimaciones se utilizaron los datos de la Encuesta Nacional de Hogares (Enaho), del INEC. El período de estudio se extiende del 2011 al 2017, lo cual garantiza la comparabilidad de los resultados. La población de estudio son las personas con edades de entre 25 y 60 años que se encontraban ocupadas en el momento de la encuesta y que reportaron un ingreso mensual mayor a cero, así como un número de horas laboradas conocido. El rango de edad se definió de esta manera debido a la tardía inserción de las mujeres al mercado laboral, por efecto de la asistencia a la educación, y se extiende hasta los 60 años porque al acercarse la edad de retiro se reduce la participación laboral femenina.

La variable dependiente corresponde al logaritmo natural del ingreso principal bruto por hora. Las Enaho registran el número de horas laboradas normalmente por semana y el ingreso principal bruto del mes anterior de las personas ocupadas. La variable dependiente se muestra en la ecuación 7.3.

ECUACIÓN 7.2

$$\Delta\bar{Y} = \underbrace{(\bar{X}_A - \bar{X}_B)'}_{\text{Dotación}} \underbrace{\beta_B}_{\text{Coeficientes}} + \bar{X}'_B \underbrace{(\beta_A - \beta_B)}_{\text{Interacción}} + \underbrace{(\bar{X}_A - \bar{X}_B)}_{\text{Interacción}} (\beta_A - \beta_B)$$

Donde:

Dotación: representa la contribución de las diferencias en las variables explicativas entre los grupos.

Coeficientes: muestra las diferencias en los coeficientes de los grupos.

Interacción: explica que las diferencias intergrupales entre variables explicativas y coeficientes pueden ocurrir simultáneamente.

ECUACIÓN 7.3

$$y = \ln \left(\frac{\text{Ingreso principal bruto total}}{\text{horas laboradas por semana} * 4,33} \right)$$

PARA MÁS INFORMACIÓN SOBRE **DISCRIMINACION SALARIAL** véase Solórzano, 2018, en www.estadonacion.or.cr

CUADRO 7.1

Variables incluidas en el modelo *probit* de condición de trabajo
(1: sí trabaja y 0: no trabaja)

Variabes	Categorías
Asistencia a educación	0. No 1. Sí
Dominio de un segundo idioma	0. No 1. Sí
Edad	Entre 25 y 60 años
Edad al cuadrado	
Escolaridad	Entre 0 y 23 años
Estado conyugal	0. Soltero(a), viudo(a), divorciado(a) 1. Casado(a) o en unión libre
Jefatura del hogar	0. No es jefe(a) de hogar 1. Es jefe(a) de hogar
Jefatura monoparental	0. No 1. Sí
Miembros menores de 6 años en el hogar	Cantidad entre 0 y 6
Región de residencia	0. Resto de regiones 1. Central
Zona de residencia	0. Rural 1. Urbana

Fuente: Solórzano, 2018.

Aportes metodológicos en materia de oportunidades, estabilidad y solvencia económicas

Estimación del ajuste fiscal requerido para estabilizar el crecimiento de la deuda pública

En 2017, el desbalance de las finanzas públicas y la acelerada expansión del endeudamiento generaron mayores presiones sobre el déficit fiscal. En este contexto, Soto (2018) estimó el ajuste necesario para estabilizar el crecimiento de la deuda pública. Cuanto mayor sea el peso de los préstamos del Gobierno con respecto al PIB, mayores serán las dificultades para cumplir con esos compromisos. El ajuste fiscal requerido depende de los cuatro factores que se detallan en la siguiente ecuación 7.4.

El primer componente es el tamaño de la deuda. Dada la falta de solvencia del Gobierno, el endeudamiento ha crecido de forma acelerada en los últimos años: de un 29% a un 49% del PIB entre 2010 y 2017. El segundo factor es la tasa de interés real, es decir, el costo del servicio de la deuda. Un aumento en este rubro incrementará el esfuerzo que debe realizar el país para sanear sus finanzas. El tercero es el balance primario (ingresos menos gastos totales, sin incluir intereses); un descenso en la recaudación y/o un incremento en el gasto primario se traducirán en un deterioro de la situación fiscal.

El último componente es el crecimiento de la economía, dado que este aproxima la disponibilidad de recursos para cumplir con los compromisos fiscales. Una desaceleración de la producción, como la registrada en 2017, aumenta la corrección requerida. El moderado crecimiento de los últimos años y la desaceleración del 2017 inciden negativamente sobre el ajuste fiscal, dado que la tasa de interés real se ha expandido más rápido que la economía.

Previo a la estimación de la brecha salarial entre hombres y mujeres fue necesario calcular un modelo de regresión probabilístico, que se empleó para corregir el sesgo de selección descrito anteriormente. Para ello la variable dependiente se dividió en dos categorías: 1 sí trabaja y 0 no trabaja. Las variables independientes se presentan en el cuadro 7.1.

Por su parte, el cuadro 7.2 detalla las variables consideradas en el modelo Oaxaca-Blinder. Se incluyeron rasgos de capital humano, como el nivel educativo y la experiencia potencial de los individuos (calculada mediante la fórmula edad menos los años de escolaridad menos 6), dado que la experiencia tiende a incrementar la productividad de los

trabajadores, lo cual justifica salarios más altos (Mincer, 1974). También se estimó la experiencia potencial al cuadrado, porque esta variable tiene rendimientos marginales decrecientes y se esperaba que tenga signo negativo.

En cuanto a las características o condiciones laborales, se incluyeron: la formalidad en el empleo, aproximada mediante el tipo de seguro en la CCSS, asumiendo que las personas aseguradas directas son formales; el sector institucional (público o privado) y la rama de actividad, agrupada en ocho categorías. Los resultados de las regresiones y de la descomposición de la técnica Oaxaca-Blinder están disponibles en la ponencia de Solórzano (2018), en el sitio www.estadonacion.or.cr.

CUADRO 7.2

Variables incluidas en el modelo Oaxaca-Blinder^{a/}

Variables	Categorías y unidad de medición
Ingreso principal por hora laborada	Colones por hora Entre 0 y menos de 100.000
Condición de aseguramiento	0. No asegurado y otras formas de seguro ^{b/} 1. Asegurado directo
Experiencia potencial	Años Entre 0 y 55
Experiencia potencial al cuadrado	
Jefatura del hogar	0. No 1. Sí
Nivel educativo	Años de escolaridad Entre 0 y 23
Posición en el empleo principal	0. Asalariado 1. Empleador 2. Cuenta propia
Rama de actividad	0. Sector agropecuario 1. Industria manufacturera 2. Comercio 3. Enseñanza y salud 4. Actividades profesionales y de apoyo administrativo 5. Servicios domésticos 6. Otras actividades sociales ^{c/} 7. Resto de actividades ^{d/}
Sector institucional	0. Público 1. Privado
Tamaño de la empresa	0. Menos de 20 empleados (pequeña) 1. 20 empleados o más (grande)
Región de residencia	0. Resto de regiones 1. Región Central
Zona de residencia	0. Rural 1. Urbana

a/ Se estimaron varios modelos, pero estas variables corresponden al que obtuvo el mejor ajuste.

b/ Otras formas de seguro: personas aseguradas mediante convenio, por el Estado, familiares de asegurados, pensionados y sus familiares, con seguro voluntario o de trabajador independiente, privado o extranjero, entre otros.

c/ Otras actividades sociales: alojamiento y servicios de alimentación, actividades artísticas y de entretenimiento, administración pública, información y comunicaciones, actividades financieras y de seguros, actividades inmobiliarias, organizaciones extra-territoriales y otras actividades de servicios.

d/ Resto de actividades: explotación de minas y canteras suministro de electricidad, gas, vapor y aire acondicionado; suministro de agua; construcción y transporte y almacenamiento.

Fuente: Solórzano, 2018.

ECUACIÓN 7.4

$$A_n = BP_n^* - BP_n$$

$$BP_n^* = D_n \times (r_n - g_n)$$

Donde:

A_n = ajuste fiscal requerido en el año 'n'.

BP_n = balance primario (ingresos menos gastos sin incluir intereses) en el año 'n'.

BP_n^* = balance primario que requiere la economía para estabilizar el crecimiento de la deuda, en el año 'n'.

D_n = tamaño de la deuda que fue adquirida en el pasado.

r_n = tasa de interés real en el año 'n'.

g_n = tasa de crecimiento real de la economía en el año 'n'.

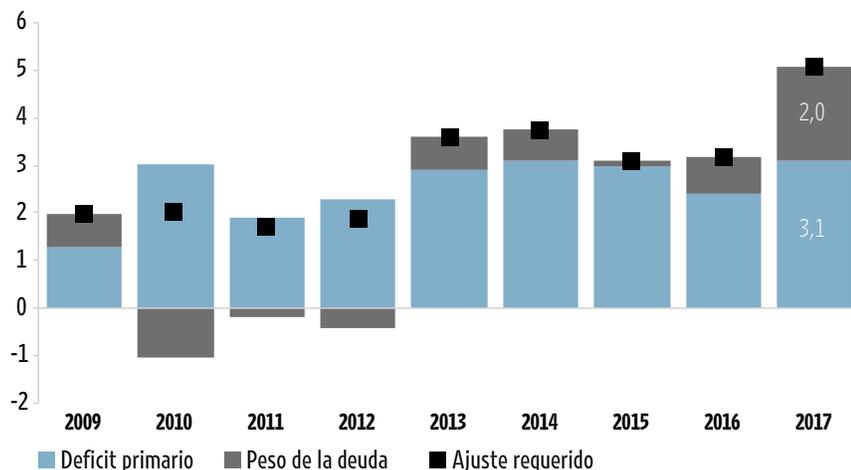
Si hubiera un superávit primario (ingresos superiores a los gastos sin incluir intereses), el ajuste sería necesario solo para hacer frente al aumento en el servicio de la deuda, el cual se mide como el efecto de las tasas de interés real sobre el tamaño de los préstamos adquiridos. Sin embargo, este no es el caso de Costa Rica, dado que actualmente existe un déficit primario de 3,1% del PIB. Esto significa que el esfuerzo debe tener dos objetivos: eliminar el déficit primario y cubrir el crecimiento del servicio de la deuda (gráfico 7.2).



PARA MÁS INFORMACIÓN SOBRE
**FINANZAS PÚBLICAS
EN COSTA RICA,**
véase Soto, 2018a, en
www.estadonacion.or.cr

GRÁFICO 7.2

Ajuste fiscal requerido para estabilizar la deuda pública y sus componentes (porcentajes)



Fuente: Soto, 2018, con datos del BCCR y el Ministerio de Hacienda.

ECUACIÓN 7.5

$$D_n^* = \frac{\sum [d_a^* \times t_a^* \times p_{n,a}]}{\sum [t_a^* \times p_{n,a}]}$$

Donde:

D_n^* = tasa de desempleo agregada en el año 'n', en el escenario de que lo único que cambió fue la demografía, mientras que todo lo demás se mantuvo al nivel de 2017.

d_a^* = tasa de desempleo del grupo de edad 'a' en 2017.

t_a^* = tasa de participación del grupo de edad 'a' en 2017. Considera a los ocupados y a las personas que buscan trabajo.

$p_{n,a}$ = proporción de personas en el grupo de edad 'a' en el año 'n'.

18 y 69 años. Además, se consideró que las tasas específicas de desempleo y participación laboral para cada uno de los grupos se mantuvieron constantes al nivel del 2017. La ecuación 7.5 detalla el cálculo realizado para proyectar la tasa de desempleo.

En cuanto a la descomposición de la tasa de desempleo, se realizó una transformación logarítmica del indicador, con el propósito de determinar el efecto de tres factores:

- Generación de empleo: medida como el número de ocupados que registran las encuestas de hogares. Una reducción en este indicador genera presiones sobre la tasa de desempleo.
- Disposición para trabajar: medida como la proporción de personas ocupadas en el mercado o que buscan empleo, con respecto al total de la población en edad de trabajar.
- Crecimiento demográfico: variación en el tamaño de la población, según grupos de edad.

La descomposición de la tasa de desempleo se muestra en la ecuación 7.6



PARA MÁS INFORMACIÓN SOBRE
**VÍNCULOS ENTRE EL CAMBIO
DEMOGRÁFICO Y EL MERCADO
LABORAL**

véase Jiménez-Fontana, 2018, en
www.estadonacion.or.cr

Determinantes de la estructura productiva que impactan la desigualdad

Meneses et al. (2018) analizaron la desigualdad en la distribución del ingreso desde la perspectiva de la estructura productiva. Para ello, las autoras realizaron una descomposición del coeficiente Gini que dimensiona cuánto de su valor total se debe a las diferencias salariales dentro y entre ramas, y cuáles brechas no provienen de una distinción según actividad económica. Para descomponer el Gini se empleó el modelo que se muestra en la ecuación 7.7.

Proyección y descomposición de la tasa de desempleo en Costa Rica

Para explorar a profundidad los vínculos entre el mercado laboral y el cambio demográfico, Jiménez-Fontana (2018) realizó un estudio que incluye una proyección de la tasa de desempleo y una descomposición de este indicador. Su principal fuente de información fueron las encuestas de hogares del INEC, cuyos factores de expansión se ajustaron por edades simples y sexo, a fin de que coincidieran con las proyecciones de población del Censo de 2011. Es importante señalar

que el cuestionario de las encuestas cambió en dos oportunidades durante el período de análisis. No obstante, para efectos de este trabajo no es posible determinar en qué medida el cambio observado obedece a esa limitación metodológica.

La proyección teórica de la tasa de desempleo se basó en el supuesto de que el único cambio fue la estructura por edades de la población. Este análisis utilizó la metodología de Fallick et al. (2010). Para la estimación se construyeron once grupos de edad para las personas de entre

ECUACIÓN 7.6

$$\log\left(\frac{E_{n+1,a}}{E_{n,a}}\right) = \log\left(\frac{O_{n+1,a}}{O_{n,a}}\right) - \left[\log\left(\frac{T_{n+1,a}}{T_{n,a}}\right) + \log\left(\frac{P_{n+1,a}}{P_{n,a}}\right)\right]$$

$$\log\left(\frac{E_{n+1,a}}{E_{n,a}}\right) \approx \frac{E_{n+1,a} - E_{n,a}}{E_{n,a}}$$

$$E_{n+1,a} - E_{n,a} = -(D_{n+1,a} - D_{n,a})$$

$$D_{n+1,a} - D_{n,a} \approx -E_{n,a} \times \left[\log\left(\frac{E_{n+1,a}}{E_{n,a}}\right)\right]$$

$$D_{n+1,a} - D_{n,a} \approx -E_{n,a} \times \left\{\log\left(\frac{O_{n+1,a}}{O_{n,a}}\right) - \left[\log\left(\frac{T_{n+1,a}}{T_{n,a}}\right) + \log\left(\frac{P_{n+1,a}}{P_{n,a}}\right)\right]\right\}$$

Donde:

$O_{n,a}$ = cantidad de personas ocupadas en el año 'n' y del grupo de edad 'a'.

$P_{n,a}$ = cantidad de personas en el año 'n' y del grupo de edad 'a'.

$T_{n,a}$ = tasa de participación en el año 'n' y del grupo de edad 'a'. Se define como la proporción de personas que están ocupadas o buscan empleo, con respecto al total de población en edad de trabajar.

$E_{n,a}$ = tasa de empleo en el año 'n' y del grupo de edad 'a'. Se define como la proporción de ocupados con respecto a la población económicamente activa.

$D_{n,a}$ = tasa de desempleo en el año 'n' y del grupo de edad 'a'. Se define como la proporción de desocupados con respecto a la población económicamente activa.

ECUACIÓN 7.7

$$I = \left(\sum_{g=1}^G \phi_g \varphi_g I_g \right) + \bar{I}_d + R$$

Donde:

ϕ_g = porcentaje de población del grupo 'g'.

φ_g = porcentaje de ingresos del grupo 'g'.

I_g = desigualdad entre grupos (cuando a cada individuo se le asigna el ingreso medio de su grupo).

\bar{I}_d = desigualdad dentro de cada grupo.

R = residuo implícito por el traslape de ingresos.

g = rama de actividad económica.

ECUACIÓN 7.8

$$G_{it} = \alpha_{it} + \beta 1 PIB_{it} + \beta 2 clave_imp_{it} + \beta 3 estrag_{it} + \beta 4 E_{it} + \beta 5 Producti_{it} + \beta 6 calif_{it} + \varepsilon_{it}$$

Donde:

G_{it} = coeficiente de Gini de la rama 'i' en el año 't'.

PIB_{it} = crecimiento promedio de los últimos cuatro años del PIB del sector 'i', en el año 't'.

$clave_imp_{it}$ = porcentaje del valor agregado producido por sectores con encadenamientos claves e impulsores de la rama 'i', en el año 2012.

$estrage_{it}$ = porcentaje del valor agregado producido por sectores con encadenamiento estratégico de la rama 'i', en el año 2012.

E_{it} = porcentaje del ingreso concentrado en el capital (excedente bruto de explotación) de la rama 'i', en el año 2012.

$Producti_{it}$ = crecimiento promedio de la productividad laboral de la rama 'i', en el periodo 2012-2017.

$calif_{it}$ = porcentaje de personas calificadas de la rama 'i', en el año 't'. Se define a una persona calificada como aquella que tiene educación secundaria completa o algún nivel mayor de instrucción.

El componente R captura este efecto, es decir, cuantifica la porción del coeficiente de Gini que se debe a efectos ajenos a la conformación de los grupos de población, como por ejemplo ubicación geográfica, sexo, edad, desempleo, educación, crisis económica, política social, entre otros.

Para aplicar esta metodología se utilizaron las encuestas de hogares del INEC. Es importante mencionar que los ajustes metodológicos implementados por esa entidad, como los cambios en los manuales de codificación para las ramas de actividad, limitan las comparaciones de largo plazo. Por ello no es posible comparar las ramas de actividad en todos los años, ni separar los efectos dentro y entre ramas. Dada esta limitación, el estudio unificó los dos componentes (dentro y entre), para tener una mejor interpretación del peso de las actividades económicas en la desigualdad, lo que permitió contar con una trayectoria completa de información para todo el período de análisis (1990-2017).

En el modelo de panel de datos se calculó, para los años 2012 a 2017, cuánto de la concentración de ingresos derivó de las características de las ramas de actividad. Además de las encuestas de hogares, se utilizó la matriz insumo-producto (BCCR, 2012). El modelo se detalla en la ecuación 7.8.



PARA MÁS INFORMACIÓN SOBRE
**DESIGUALDAD Y ESTRUCTURA
PRODUCTIVA**

véase Meneses et al., 2018, en
www.estadonacion.or.cr

**Aportes metodológicos en materia de
transporte y movilidad**

**Estimación de los tiempos de viaje de
las personas trabajadoras en la GAM**

Para estimar los tiempos de viaje en la Gran Área Metropolitana (GAM), de las personas que trabajan en un cantón distinto al de su residencia, Sánchez (2018b) utilizó la plataforma digital *Google Maps*. Esta es una aplicación que, por medio de los teléfonos celulares, guía a los usuarios en sus trayectos y registra información sobre sus movimientos. Con esta herramienta, y mediante el uso de sistemas de información geográfica (SIG), se construyó una base de datos sobre tramos recorridos, rutas usadas, horarios, días, velocidades y cantones involucrados.

Los tiempos de viaje estimados por *Google Maps* están compuestos por varios factores, que dependen de los datos disponibles para cada área particular. Estos incluyen los límites de velocidad oficiales y recomendados, las velocidades basadas en el tipo de carretera, velocidades medias históricas en ciertos períodos (o en momentos específicos del día), tiempos efectivos de viaje de otros usuarios e información del tráfico en tiempo real. Todos estos elementos forman parte del algoritmo que utiliza Google para calcular el tiempo de recorrido en diferentes tramos de una ruta. Google Maps gestiona el tráfico en vivo, comparando las predicciones con los tiempos reales, para modificar sus algoritmos y fuentes de datos, y mejorar los pronósticos en términos medios.

Al combinar la velocidad de un usuario con la de otros que circulan en la misma carretera, cruzándola con la información de miles de teléfonos celulares que se mueven por una ciudad a cualquier hora, *Google Maps* puede conseguir un esquema muy preciso, y en directo, de

las condiciones del tráfico. Es importante mencionar que esta información no se encuentra disponible en una base de datos, lo cual implica que, en el ejercicio realizado para este Informe, para cada tramo y ruta que conecta los cantones de la GAM, en cada horario y día seleccionado, fue necesario obtenerla en el sitio <https://www.google.es/maps>, para ir conformando la base de datos mediante el uso de SIG, tramo por tramo.

Una vez sistematizada la información, se procedió a estimar los tiempos de viaje en momentos de flujo sin y con congestión. En el primer caso, se trata del tiempo de recorrido en la red vial cantonal que conecta los municipios, en dos períodos del día: diurno (9 am a 3 pm) y nocturno (8 pm a 4 am). En el segundo, dado que los valores de la variable “velocidad” cambian (disminuyen), por la mayor cantidad de vehículos en circulación, fue preciso reestimar los tiempos. Así, los intervalos de congestión que se consideraron fueron: de 5 a 9 am y de 4 a 8 pm.

En ambos casos los datos analizados correspondieron a dos grupos de observaciones: el primero incluye los meses de enero, marzo, mayo, julio, septiembre y noviembre de 2016, y el segundo los meses de febrero, abril, junio, agosto, octubre y diciembre de 2017. En 2016, para cada mes, se seleccionaron las semanas 2 y 4 (de lunes a viernes) y en 2017 las semanas 1 y 3. Los datos se promediaron por hora, día, mes y año. En total se monitorearon 12 meses, 52 semanas, 260 días y 2.080 horas. Las matrices de origen-destino derivadas de este ejercicio se pueden consultar en los anexos de la ponencia de Sánchez (2018b) en el sitio www.estadonacion.or.cr.

**Estimación de los costos derivados
del congestionamiento vial para
las personas que se movilizan de
un cantón a otro de la GAM, para
trabajar**

Con base en las estimaciones de tiempo de viaje antes descritas, Sánchez (2018b) calculó los costos que ocasiona el congestionamiento vial a las personas trabajadoras de la GAM que se movilizan de un cantón a otro. Para esto utilizó una matriz de origen/destino (lugar de

residencia/lugar de trabajo), basada en el Censo de Población 2011 del INEC (el último disponible) e identificó la cantidad de empleos que se generan en cada cantón, así como el número de trabajadores que viven en el mismo municipio y trabajan en otro (Arias y Sánchez, 2012). Dado que el Censo no mide ingresos, se usó la Encuesta Nacional de Hogares (2017) para asignar un monto promedio a cada persona trabajadora, según su ocupación.

Para cada actividad productiva se estimó una matriz, considerando las siguientes ocupaciones: i) nivel directivo, administración pública y empresa privada, ii) nivel profesional científico e intelectual, iii) nivel técnico y profesional medio, iv) apoyo administrativo, v) ventas locales y prestación de servicios directos, vi) actividades agropecuarias y pesqueras calificadas, vii) producción artesanal, construcción mecánica, artes gráficas, viii) operación de instalaciones, máquinas y ensambladores y ix) ocupaciones elementales.

Con la información recopilada, se determinó el ingreso por hora de los ocupados en las distintas actividades productivas y se estimaron las horas en tiempos de viaje que invierte cada uno, por ocupación, y para cada destino al que se moviliza. Posteriormente se obtuvo la diferencia entre los tiempos de viaje en flujos con y sin congestionamientos, y sobre esa diferencia se valoró el costo económico en que incurren las personas por el tiempo perdido en los atascos. Es decir, se midió: el tiempo de viaje en horas perdidas en congestionamientos por los trabajadores que laboran fuera del cantón donde viven, el salario por hora que devengan en promedio según su ocupación, y el costo monetario final en que incurren debido a ese retraso en su tiempo. Los costos se tasaron en millones de dólares al año. Por último, se clasificaron los 31 cantones de la GAM en un *ranking* que muestra los lugares en los que las personas trabajadoras incurren en los mayores costos, tanto absolutos como per cápita.



PARA MÁS INFORMACIÓN SOBRE
**SITUACIÓN GENERAL
DEL TRANSPORTE**

véase Sánchez, 2018b, en
www.estadonacion.or.cr

**Análisis espacial sobre la congestión
vehicular a partir de la base de datos
de la aplicación Waze**

Existen nuevas fuentes de información que apenas comienzan a aprovecharse y que permiten conocer la magnitud, ubicación y características de los congestionamientos viales. El PEN realizó una primera investigación explorando la base de datos de la aplicación móvil Waze, a partir de la información sobre atascos para los años 2017 y 2018 (Durán y León, 2018)⁵. El primer análisis utilizó los datos de atascos (*jams*) en 2017, para describir el comportamiento general de las zonas de mayor congestión vehicular en Costa Rica. La base de datos consta de poco menos de 25 millones de registros, con la información que Waze registra diariamente y traslada al MOPT, organizada en trece variables:

- *City*: ciudad donde se registra el atasco.
- *Delay*: retraso en segundos con respecto a la velocidad promedio del segmento en vía libre.
- *EndNode*: punto donde termina el atasco (referencia geográfica).
- *Length*: longitud del atasco en metros.
- *Line*: lista de pares de coordenadas (longitud y latitud) a lo largo de cada atasco.
- *RoadType*: tipo de camino en que se reporta el atasco.

- *Speed*: velocidad promedio registrada en el atasco en metros por segundo.
- *StartNode*: punto donde inicia el atasco (referencia geográfica).
- *Street*: nombre o número de calle donde se registra el atasco (corresponde a los segmentos definidos por Waze para Costa Rica).
- *Año*: 2017.
- *Mes*: mes del año en que se registran los datos.
- *Día*: fecha del evento.
- *Hora*: hora del evento (formato de 24 horas).

Posterior a la carga de los datos, se analizó el comportamiento, tipo y estructura de cada una de las variables. En su mayoría, estas son de tipo numérico o texto; sin embargo, en el caso de *line* corresponde a una lista de pares de coordenadas que indican la ubicación y extensión territorial de cada congestionamiento. En forma paralela, se verificó la existencia de vacíos o inconsistencias de información y se eliminaron las variables que reportaron un alto porcentaje de valores faltantes (*city*, *endNode*, *startNode*). Además, se excluyó la variable *roadType*, pues contiene categorías asignadas por los desarrolladores de Waze que no brindaban información útil para los propósitos de este análisis. En la variable *delay*, o retraso, se identificaron valores negativos que no se consideraron en esta medición. Con estas correcciones la base se redujo a poco más de 19 millones de atascos.

Luego se procedió a extraer la información del conjunto de coordenadas para ubicar espacialmente cada *jam*, y se incorporaron los polígonos de los distritos de Costa Rica en 2018, a fin de contar con un referente geográfico que permitiera

sustituir la variable *street*, que contenía gran cantidad de nombres repetidos. De esta forma, para cada atasco se tenía la ubicación georeferenciada, la provincia, el cantón y el distrito donde ocurrió. Esto permitió agregar la cantidad de *jams* ocurridos durante el 2017 en cada distrito y clasificarlos en cuatro categorías según el nivel de congestión vial en: i) muy alto, ii) alto, iii) medio y iv) bajo.

El análisis se concentró en las zonas con muy alto nivel de congestionamiento. Se trabajó con 28 distritos ubicados en el centro del país y 10 millones de registros. El análisis se realizó en tres etapas. La primera consistió en una caracterización general de los atascos en los distritos prioritarios: velocidad promedio, tiempos de retraso, extensión del atasco y distribución de los eventos a lo largo del año.

En la segunda etapa se agregaron dos variables, que aportaron información sobre: i) densidad poblacional y ii) porcentaje de personas que se desplazan del cantón diariamente para trabajar. Ambos datos provienen del INEC (Censo 2011). Esto permitió aumentar la capacidad explicativa de la base y hacer una mejor caracterización de las posibles causas del congestionamiento en la GAM.

El tercer paso consistió en identificar conglomerados de atascos según su comportamiento en las variables incluidas en el análisis: *length*, *delay*, *speed*, densidad poblacional, porcentaje de personas que se desplazan, día, mes y hora en que ocurrió el atasco. Se utilizó la técnica de análisis multivariado llamada K-medias, con 1.000 iteraciones y 30 repeticiones para determinar la cantidad de conglomerados, y con 1.000 iteraciones y 50 repeticiones para la selección del algoritmo *Lloyd*. Seguidamente se estudiaron los tres conglomerados resultantes y su distribución espacial en el territorio nacional. Los principales hallazgos se detallan en el capítulo 6 de este Informe.

Este Anexo Metodológico fue preparado por Karen Chacón, Pamela Jiménez, Leonardo Merino y Natalia Morales.

NOTAS

- 1** El valor muestra el grado de concentración de la actividad turística en el distrito, en relación con el resto del país. Cifras mayores a 1 indican una concentración más que proporcional en ese distrito. El propósito de incluir este indicador es identificar cuáles localidades tienen fuertes concentraciones de esta actividad y si esta se asocia a los homicidios (Sánchez, 2018a).
- 2** En Estadística se dice que existe heterocedasticidad cuando los errores no son constantes a lo largo de toda la muestra. Este problema tiene errores en los cálculos del estimador de la matriz de varianzas y covarianzas, así como en la pérdida de eficiencia en los estimadores, al no tener varianza mínima.
- 3** En Estadística la multicolinealidad ocurre cuando algunas variables predictoras incluidas en el modelo están correlacionadas con otras variables predictoras. Este problema incrementa la varianza de los coeficientes de regresión, haciéndolos inestables y poco precisos.
- 4** La regresión logística es un tipo de análisis utilizado para predecir el resultado de una variable categórica (que puede adoptar un número limitado de categorías) en función de las variables independientes o predictoras. Es útil para modelar la probabilidad de que ocurra un evento en función de otros factores.
- 5** Este estudio se efectuó gracias a una alianza institucional con el Departamento de Informática del MOPT, el ente encargado de realizar la descarga de los datos para Costa Rica. Además, se contó con el apoyo técnico y logístico del Laboratorio Nacional de Computación Avanzada (CNCA) del CeNAT-Conare, para almacenar y fusionar los archivos en una misma base de datos.