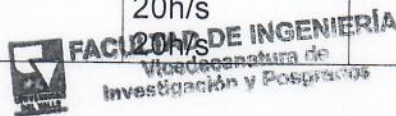


Fecha de presentación del Informe: Día Mes Año **1. Datos generales del Proyecto**

Código del proyecto: CI 2930			
Título del proyecto: Comportamiento de las estimaciones de los parámetros de desempeño de criterios de clasificación: un caso de aplicación en diagnóstico de enfermedades			
Facultad o Instituto Académico: Facultad de Ingenierías			
Departamento o Escuela: Escuela de Estadística			
Grupo (s) de investigación: INFERIR			
Entidades: Universidad del Valle			
Palabras claves: Diagnóstico, procesos de clasificación, tasa de correctos clasificados, tasas de mal clasificados			
Investigadores ¹	Nombre	Tiempo asignado	Tiempo dedicado
Investigador Principal	José Rafael Tovar Cuevas	10h/s	
Coinvestigadores	Lyda Osorio	5h/s	
	Luis Carlos Bravo	20h/s	
	Jennyfer Portilla	25h/s	
	Diosimar Cardoza	25h/s	
Otros participantes	Andrés Camilo Mendez	20h/s	
	Angie Castro	10h/s	
	Juliette Barbosa	10h/s	
	Michael Steven Rodríguez	10h/s	
	Juan David Díaz Mutis	20h/s	
	Jhon Edinson López	20h/s	



05 FEB 2018

RECIBIDO

¹ Todas las personas relacionadas en el informe y que participen en el proyecto deben haber suscrito el acta de propiedad intelectual de acuerdo con los formatos establecidos.

2. Resumen ejecutivo:

En el proyecto se desarrollaron diferentes propuestas de reglas y procedimientos que pueden ser utilizados para clasificar individuos que presentan alguna característica de salud de interés (infección, enfermedad, discapacidad etc). Se desarrolló metodología estadística apoyada en la teoría de la probabilidad para obtener procedimientos que permitan identificar individuos que presenten algún evento de interés en salud. Las metodologías estadísticas utilizadas incluyeron el trabajo con variables latentes y construcción de indicadores y el uso de metodología bayesiana para la construir, validar y evaluar algoritmos que sirvan como reglas de clasificación. Se estimaron los parámetros de desempeño (sensibilidad, la especificidad y el área bajo la curva ROC), en situaciones en las que, la expresión de la condición de salud, se observa a través de biomarcadores o características no observables (variables latentes) cuya representación matemática se asume cuantitativa continua en escala de intervalos o de razón. Los modelos planteados incluyeron la presencia de dependencia estadística entre las variables de interés. Se ilustra la aplicación de los procedimientos y métodos utilizando datos de individuos con dengue e individuos con síndrome de frágil X.

3. Síntesis del proyecto:

Objetivo General

Evaluar el comportamiento estadístico de las estimaciones de los parámetros de desempeño y la prevalencia en estudios de validación de pruebas para diagnóstico clínico con respuesta cuantitativa continua y dependencia tipo cópula usando métodos convencionales y una metodología alternativa.

Objetivos Específicos

1. Estimar los parámetros de desempeño de las pruebas de respuesta continua utilizando curvas características de operación (Curvas ROC), empleando procedimientos de inferencia clásica y métodos propios de la inferencia bayesiana.
2. Estimar los parámetros de desempeño y la prevalencia utilizando la reparametrización introducida por Tovar y Achcar (2011) empleando procedimientos propios del paradigma frecuentista y el paradigma bayesiano.

3. Estudiar el efecto de la dependencia lineal débil y no lineal sobre las estimaciones obtenidas mediante los dos procedimientos y en los puntos de corte establecidos al usar las curvas ROC.
4. Estimar los parámetros de desempeño de las pruebas de respuesta continua con estructura de dependencia tipo cópula FGM y Gumbel utilizando curvas características de operación (curvas ROC) para datos dependientes.
5. Aplicar los métodos estadísticos desarrollados en los objetivos específicos previos con los datos obtenidos en un estudio de cohorte multicéntrica de dengue realizada previamente en Colombia.

Metodología

Para estudiar el efecto de la dependencia entre variables, el modelamiento estadístico consideró incluir parámetros tipo copula. En un modelo se incluyó la función copula FGM que modela dependencias linealmente débiles y en el otro se incluyó la estructura de dependencia tipo función copula de Gumbel Barnett, la cual, modela dependencias débiles pero no necesariamente lineales. Se estudió el efecto de la dependencia sobre las estimaciones de los parámetros de desempeño y el área bajo la curva ROC. El modelamiento incluyó una reparametrización de los parámetros de la distribución multinomial asociada al modelo de muestreo con el que se obtienen los datos en campo cuando se realizan estudios de validación pruebas para diagnóstico clínico. Adicionalmente, se estimaron las áreas bajo las curvas ROC usando métodos convencionales que asumen normalidad bivariada con independencia en la distribución de los biomarcadores y usando una nueva metodología (propuesta por el investigador del proyecto) que permite realizar las estimaciones sin tener que considerar la normalidad e incluye en el modelo un parámetro de dependencia tipo copula. Se realizó una comparación entre los resultados obtenidos con ambas metodologías a través de estudios de simulación y con los datos de un grupo de individuos con diagnóstico positivo para dengue.

También se desarrolló y validó un índice que permitiera clasificar individuos de acuerdo con el grado de discapacidad a partir de la pregunta 20 formulada en el cuestionario "Genética Médica Poblacional" aplicado a la población en estudio. Para el desarrollo del índice se utilizaron 3 técnicas: proceso de jerarquía analítica, análisis de componentes

principales y análisis factorial múltiple. Los puntos de corte fueron obtenidos mediante la curva ROC y valores mínimos y máximos de grupos formados para los niveles de discapacidad. La selección del mejor índice se llevó a cabo por medio del análisis de sensibilidad, especificidad y la concordancia por medio del índice de Kappa ponderado.

Finalmente, para la estimación del área bajo la curva ROC en pruebas para diagnóstico clínico con dependencia tipo cópula se propone un método alternativo que combina dos biomarcadores continuos para estimar el área bajo la curva ROC (AUC) para clasificación de individuos en un estudio de tamizaje, considerando el efecto de la dependencia entre los resultados de dos pruebas al modelarla con estructura tipo cópula. Se asumió presencia de estructuras de dependencia lineales débiles (cópula FGM) y estructuras de dependencia muy débil pero no necesariamente lineal (cópula Gumbel- Barnett). Nuestro método, al proponer la función cópula como medio para llegar a la estimación del AUC, suprime la limitación de asumir específicamente una distribución normal para los datos.

Resultados obtenidos

Basados en los resultados obtenidos bajo el supuesto de dependencia en un modelo binormal, se observó que el comportamiento de las estimaciones bajo el modelo que asume dependencia cópula FGM (reparametrización), presenta un comportamiento bastante adecuado, ya que los valores del Error Cuadrático Medio (ECM) se reducen y se presenta un aumento en la cobertura de los intervalos estimados, esto cuando se compara con los otros dos modelos implementados. Adicionalmente, se observó que realizando el ajuste por medio de las curvas ROC obtenidas usando el método propuesto que no considera el supuesto de binormalidad e incluye la dependencia, se mejoran las estimaciones de las especificidades y una de las sensibilidades, pues se tienen menores ECM y aumenta la cobertura, esto especialmente cuando se consideran niveles de dependencia moderada y fuerte. En el estudio usando los datos de los niveles de leucopenia y trombocitopenia obtenidos en una cohorte de individuos diagnosticados con dengue, se observó que la estimación del área bajo la curva (AUC) cuando se asumió una estructura de dependencia tipo copula FGM fue del 94%.

Respecto al índice para realizar clasificación de individuos por niveles de discapacidad, se compararon dos métodos, el de jerarquía analítica y un método desarrollado combinando procedimientos estadísticos multivariados y heurística. Se observó que el índice con mejor desempeño fue el obtenido por medio del proceso de jerarquía

analítica, simulando puntajes a través de la distribución beta y usando como prueva de referencia la prueba genética. Este índice presenta una sensibilidad de 80% para discapacidad y una especificidad de 97 %.

Principales conclusiones y/o recomendaciones

El comportamiento de las estimaciones obtenidas usando la reparametrización propuesta por Tovar & Achcar (2012) se ajusta bastante bien a lo esperado, pues presenta un error cuadrático medio (ECM) bajo y las regiones de credibilidad además de ser muy estrechas, presentan una cobertura alta. El método propuesto para estimar el área bajo la curva ROC considerando la dependencia tipo copula, mostró un mejor desempeño en las estimaciones (menor ECM y mayores probabilidades de cobertura), esto con referencia al modelo de curva ROC asumiendo binormalidad y a la metodología convencional de la curvas características de operación bajo independencia, implicando que dicho modelo es adecuado para la estimación de los parámetros de desempeño clínico cuando se estudia la validación de pruebas para diagnóstico clínico que se presentan en naturaleza continua cualitativa y tienen presencia de dependencia lineal débil.

En cuanto a las metodologías implementadas para la construcción de los índices (PJA, ACP y AFM), se puede concluir que las tres permitieron una buena clasificación de los individuos. El índice construido a partir de la opción 2 del proceso de jerarquía analítica usando como referencia la prueba genética, permitió los mejores resultados en la clasificación de las personas con problemas de discapacidad. Es oportuno rescatar que el uso del proceso de jerarquía analítica, el cual es considerado heurístico o empírico, en este caso resultó ser preciso; su metodología aportó consistencia a los pesos del índice, ya que no se le asignó el mismo peso a cada uno de los ítems dentro de cada categoría de discapacidad. Con los resultados obtenidos, se demuestra la importancia del uso de herramientas estadísticas en el campo de la salud, ya que permite un acercamiento a la medición de las variables latentes, las cuales son de gran interés en dicha línea de investigación.

Cuando se efectúan estudios de validación de procedimientos para realizar diagnóstico clínico es posible contar con dos o más pruebas que permitan clasificar a los sujetos de acuerdo con su estado de salud. Dado que las pruebas están siendo aplicadas al mismo sujeto, es posible observar dependencias de alguna clase entre los resultados.

Sin embargo, la dependencia existente no necesariamente puede ser fácilmente identificable o ser de tipo lineal. Se evidenció que el método propuesto permite obtener resultados dentro de lo esperado, puesto que cuando se comparó con un método reportado en la literatura no se obtuvieron resultados tan diferentes, y de hecho se consiguieron valores más altos en las estimaciones, además una vez logra detectar la dependencia entre los biomarcadores funciona de manera adecuada.

4. Impactos actual o potencial:

La investigación en salud pública, es un tema actual y de interés para la toma de decisiones y la implementación de procesos clínicos que permitan identificar la presencia, desarrollo y control de factores que condicionan el estado de salud colectivo. Un aspecto de interés dentro de las investigaciones en salud es el relacionado con el desarrollo y validación de procedimientos biológicos que permiten la clasificación correcta de individuos según su estado real de salud. Con los resultados obtenidos, se demuestra la importancia del uso de herramientas estadísticas en el campo de la salud.

Por otro lado, esta investigación contribuyo en el ámbito académico, aportando a la formación de 7 estudiantes de pregrado y 2 estudiantes de maestría.

5. Productos:

Tabla No. 1. Cantidad y tipo de productos pactados en el Acta de Trabajo y Compromiso y productos finalmente presentados

TIPO DE PRODUCTOS	No. de PRODUCTOS PACTADOS				No. de PRODUCTOS PRESENTADOS			
Productos de nuevos conocimientos								
Artículo completo sometido en revistas indexadas	A1	A2	B	C	A1	A2	B	C
	1	2			1	2		
Formación de recursos humanos	No. de estudiantes vinculados		No. de tesis		No. De estudiantes Vinculados		No. De tesis	
Estudiantes de pregrado					3		2	

TIPO DE PRODUCTOS	No. de PRODUCTOS PACTADOS		No. de PRODUCTOS PRESENTADOS	
Semillero de Investigación			4	4
Estudiantes de maestría			2	2
Estudiantes de doctorado				
Joven investigador				
Productos de divulgación				
Ponencias presentadas en eventos (congresos, seminarios, coloquios, foros)	No. de ponencias nacionales	No. de ponencias internacionales	No. de ponencias nacionales	No. de ponencias internacionales
Propuesta de investigación			2	

Tabla No. 2. Detalle de productos

Para cada uno de los productos obtenidos y relacionados en la tabla anterior, indique la información solicitada para cada uno, anexando copia de las respectivas constancias. Como anexo a este formato encontrará el instructivo para instructivo para la revisión de informes finales y productos

Tipo de producto:	Artículo.
Nombre General:	Revista Journal of Biopharmaceutical Statistics.
Nombre Particular:	Artículo Cutpoint selection for discretizing continuos diagnostic tests with limited response
Participantes:	Jose Rafael Tovar Cuevas, Jorge Alberto Achcar, Diosimar Cardoza, Jennyfer Portilla

Tipo de producto:	Artículo.
Nombre General:	Revista Computational Statistics and Data Analysis

Nombre Particular:	Artículo Estimation of the AUC in clinical diagnostic test with copula-type dependence
Participantes:	Luis Carlos Bravo Melo, Jennyfer Portilla, Jose Rafael Tovar

Tipo de producto:	Artículo.
Nombre General:	Revista Biomédica: Revista del Instituto Nacional de Salud
Nombre Particular:	Artículo Desarrollo de algoritmos clínicos para el diagnóstico de dengue en Colombia.
Participantes:	Diana María Caicedo, Lyda Osorio, Andrés Camilo Méndez, José Rafael Tovar, Jairo Andrés Celis, Liliana Villegas, Constanza Collazos

Tipo de producto:	Tesis de pregrado
Nombre Particular:	Estimación de los parámetros de desempeño en estudios de validación de pruebas para diagnóstico clínico; cuando los trazos médicos se expresan en forma continua y presentan dependencia tipo cópula FGM.
Participantes:	Diosimar Cardoza Torres

Tipo de producto:	Tesis de pregrado
Nombre Particular:	Desarrollo y validación de un índice que permita clasificar individuos con problemas de discapacidad a partir de la pregunta 20 formulada en el cuestionario: Genética médica poblacional.
Participantes:	Jhon Edinson López León y Juan David Díaz Mutis

Tipo de producto:	Tesis de maestría
Nombre Particular:	Estimación del Área Bajo la Curva ROC en pruebas para diagnóstico clínico con dependencia tipo copula.
Participantes:	Luis Carlos Bravo Melo.



Tipo de producto:	Tesis de maestría
Nombre Particular:	Estimación del parámetro de dependencia tipo copula para variables continuas, usando datos binarios obtenidos en procesos de clasificación.
Participantes:	Jennyfer Portilla Yela

Tipo de producto:	Semillero de investigación
Nombre Particular:	Escritura de artículo para revista médica.
Participantes:	Andrés Camilo Méndez Alzate
Formas organizativas:	INFERIR

Tipo de producto:	Semillero de investigación
Nombre Particular:	Realización de artículo de revisión sobre métodos desarrollados para estimar parámetros en estudios de validación de pruebas para diagnóstico clínico.
Participantes:	Angie Castro Rengifo
Formas organizativas:	INFERIR

Tipo de producto:	Semillero de investigación
Nombre Particular:	Realización de un artículo para someter a revista biomédica.
Participantes:	Juliette Barboza Cándelo
Formas organizativas:	INFERIR

Tipo de producto:	Semillero de investigación
-------------------	----------------------------



Nombre Particular:	Escritura de artículo de revisión y nueva propuesta de investigación.
Participantes:	Michael Steven Rodríguez Cáceres
Formas organizativas:	INFERIR

Tipo de producto:	Evento de divulgación
Nombre General:	XV encuentro científico 100 años construyendo ciencia y salud
Nombre Particular:	100 años construyendo ciencia y salud
Ciudad y fechas:	Bogotá Colombia, noviembre 20, 21 y 22 de 2017
Participantes:	Diana María Caicedo, Andrés Camilo Méndez, José Rafael Tovar, Lyda Osorio
Formas organizativas:	Instituto Nacional de Salud

Tipo de producto:	Evento de divulgación
Nombre General:	XXVII Simposio Internacional de Estadística
Nombre Particular:	Estadística Aplicada
Ciudad y fechas:	Medellin Colombia, agosto 8-12 de 2017
Participantes:	Jennyfer Portilla Yela
Formas organizativas:	Universidad Nacional de Colombia



La presente versión del informe contiene las observaciones de los evaluadores:

Firma del investigador principal

VoBo. Vicedecano de Investigaciones

Por favor presente su informe impreso y en formato digital en hoja tamaño carta, letra arial 11, con espacios de 1 1/2