Uso del Modelo Lineal Generalizado *(clolog)* para la predicción del SXF en sus formas: Mutación completa, Premutación y Zona Gris

Mónica Montaño Hurtado

Universidad del Valle
Facultad de Ingeniería, Escuela de Estadística
Santiago de Cali, Colombia
2017

Uso del Modelo Lineal Generalizado *(clolog)* para la predicción del SXF en sus formas: Mutación completa, Premutación y Zona Gris

Mónica Montaño Hurtado

Trabajo de grado presentado como requisito parcial para optar al título de:
Estadístico

Director:
Ph.D José Rafael Tovar Cuevas

Universidad del Valle
Facultad de Ingeniería, Escuela de Estadística
Santiago de Cali, Colombia
2017

Dedicatoria

A Dios, por la vida y las capacidades otorgadas para lograr mis objetivos. A mis padres Porfiria Montaño y José Gil Montaño, quienes con sus esfuerzos me han dado lo que no han tenido para brindarme mejor calidad de vida por medio de la educación. A la familia Buelvas quienes me aceptaron en sus vidas e inculcaron competencias que han sido de gran importancia para el cumplimiento de mis metas. A Christian Arvey Riascos por su confianza y apoyo incondicional que me fortalecieron para culminar mi carrera profesional. A mis compañeros, en especial a mi grupo de apoyo académico compuesto por Juliette Barboza, Yesica Bejarano, Juliana Jáuregui y Camila Torres quienes estuvieron presentes de forma incondicional durante mi formación profesional y personal.

Agradecimientos

Al profesor José Rafael Tovar Cuevas, Ph.D en Estadística, profesor de la escuela de estadística de la Universidad del Valle por aceptar dirigir mi trabajo de grado y proveer en mi formación profesional y para mi desarrollo, sus conocimientos. A todos los profesores de la escuela de estadística de la Universidad del Valle, partícipes de mi formación profesional el cual fundaron en mi el pensamiento estadístico, a mis compañeros por el apoyo y moldear para mi vida competencias como trabajo en equipo tolerancia entre otras, a mi familia y amigos por alentarme y ayudarme a persistir en mis objetivos trazados.

Resumen

Este estudio explora las relaciones de las variables: antecedentes familiares con Síndrome de X Frágil (SXF), dos índices de discapacidad construidos en un estudio previo a través de los resultados de la pregunta 20 de la encuesta "Genética Medica Poblacional" y el índice de discapacidad obtenido de la aplicación del test WHODAS 2.0, con la presencia o ausencia del (SXF) en sus formas Mutación completa, Premutación y Zona gris, por medio de los modelos lineales generalizados para datos binarios con enlace log log complementario.
Las variables incluidas en los modelos planteados para explicación del trastorno Mutación Completa y Premutación, fueron significativos, a diferencia de los modelos cloglog para el trastorno Zona gris, ya que es un trastorno que no asocia problemas de discapacidad. Por otro lado, las medidas de exactitud obtenidas de la aplicación de la curva ROC evidenciaron mejor poder discriminante para el modelo ajustado para dar predicción del trastorno Mutación Completa que aquel ajustado para dar predicción del trastorno Premutación.
Para ámbitos aplicativos, se sugirió usar los modelos donde en su conjunto de variables explicativas se encuentra el índice de discapacidad obtenido del test WHODAS 2.0, ya que gracias a sus buenas características y a su fácil obtención se vuelve el método más práctico para su replicación.
Palabras claves: SXF, Discapacidad, Modelo lineal generalizado, Curva ROC

Abstract

This study explores the relationships of the variables: family history with Fragile X Syndrome (SXF), two indexes of disability constructed in a previous study through the results of question 20 of the survey "Population Genetic Medicin" and the index of disability obtained from the application of the WHODAS 2.0 test, with the presence or absence of (SXF) in its forms of Complete Mutation, Premutation and Gray Zone, using the generalized linear models for binary data with complementary log log link.
The variables included in the models proposed to explain the disorder Complete mutation and premutation were significant, unlike the models of block registration for the ZG disorder, since it is a disorder that does not associate disability problems. On the other hand, the precision measurements obtained from the application of the ROC curve showed a better discriminant power for the adjusted model to predict the Mutation Complete than the adjusted to predict the Premutation disorder.
For the application areas, it was suggested to use the models where the set of explanatory variables is the disability index obtained from the WHODAS 2.0 test, since thanks to its good characteristics and easy obtaining it becomes the most practical method for its replication.
small Keywords: SXF, disability, generalized linear model, ROC curve

Contenido

Lista de Figuras

Lista de Tablas

1. Introducción

Este estudio presenta el desarrollo de una propuesta metodológica para modelar la presencia del trastorno Mutación Completa (MC), Premutación (PM) y Zona Gris (ZG), y permitir dar predicción de ellas. La información con la que se desarrolló este trabajo fue con base al proyecto de investigación de profesionales médicos de la universidad UC Davis y la Universidad del Valle, en el cual por medio de pruebas genéticas para el SXF confirmaron las causas de discapacidad observadas por largo tiempo. Además de la prueba genética, los investigadores aplicaron el cuestionario "Genética Médica Poblacional" sobre gran parte de la población, para dar información de los antecedentes, condiciones de vida y estado salud de quienes lo diligenciaron. Sin embargo, como sucede en diferentes investigaciones, no todas las personas que presentaron resultado de la prueba genética realizaron la encuesta y en sentido inverso. En tal motivo, para el desarrollo del estudio previo de Lopez & Diaz (2017), solo tomaron en cuenta las personas que hubiesen contestado el cuestionario "Genética Médica Poblacional' y presentado los resultados de la prueba genética, para la construcción de dos índices por dos metodologías diferentes con el objetivo de dar medición de la variable latente discapacidad. Estos índices fueron construidos a partir de los resultados de la pregunta 20 del cuestionario y para la validación de ellos se hizo uso del test WHODAS 2.0, aplicado a los individuos con resultados del cuestionario y prueba genética por medio de una metodología de muestro sobre ellos.

De este modo, la variable latente discapacidad medida por medio de los índices construidos en el estudio previo y del test WHODAS 2.0, los resultados de la pregunta 18 y 19 del cuestionario "Genética Medica Poblacional", seleccionadas según criterio del experto, fueron incluidas en los modelos lineales generalizados para datos binarios con enlace log log complementario propuestos para modelar y predecir la probabilidad que un individuo presente MC, PM o ZG. Luego por medio de las curvas ROC y sus medidas de exactitud se evaluó la capacidad predictiva de los modelos y por medio de la técnica de remuestreo bootstrap se valoró la capacidad clasificatoria de los modelos al ingresar otra información muestral. Finalmente, se ajustaron cuatro diferentes modelos para modelar cada trastorno en función de las variables descritas. En el caso de la modelación del trastorno ZG, ya que las personas con este tipo de trastorno no evidencian problemas de discapacidad y por ende los instrumentos de medición no dan evidencia de ello, las variables incluidas en este modelo no fueron significativas y el modelo no pudo dar predicción del trastorno ZG. Para los trastornos MC y PM, la combinación de algunas variables explicativas evidenció lo denominado separación estadística, limitando el proceso de estimación, por dicho motivo

estos modelos fueron eludidos. Finalmente, aquellos modelos con combinación de covariables que presentaron un comportamiento dentro de lo esperado son los reportados en este estudio.

2. Formulación del problema

El Síndrome X Frágil (SXF) es un trastorno de causa genética hereditaria ligada al cromosoma X que produce retraso mental o discapacidad intelectual de severidad variable (leve, moderado, hasta grave), que puede presentarse tanto en hombres como en mujeres (Minguez 2006). Sin embargo, Lorente (2017) señala que generalmente los hombres son más afectados que las mujeres, es así como para el caso de las mujeres las manifestaciones no son valoradas con facilidad y por lo tanto no se diagnostica el propio síndrome (Martín 2015).

El trastorno puede manifestarse con alteraciones oculares como el estrabismo, que se presenta en un 35-50 % de los casos, siendo éste el más frecuente según Minguez (2006). Además, señala que los niños con SXF presentan frecuentes alteraciones auditivas, como otitis de repetición. En cuanto a los aspectos cognitivos y conductuales, presentan retraso psicomotor, hiperactividad y falta de atención, trastornos del lenguaje, la lectura y la escritura, rasgos autistas como mantenimiento escaso de la mirada, timidez, entre otras (Martorell 2010).

En cuanto a la cuantificación de este trastorno, posterior a 1991 del descubrimiento del gen FMR1 y la prueba diagnóstica para el SXF, la estimación de la prevalencia en hombres fue de 1/4.000, desde entonces a partir de los diferentes estudios se ha estimado un rango para la prevalencia del SXF en hombres de 1 en 3.717 a 1 en 8.918 (Coffee et al. 2009), mientras que en las mujeres se ha estimado una prevalencia de 1/6.000 (Mingarro Castillo et al. 2017).

En un pequeño corregimiento del municipio de Bolívar, del departamento del Valle del Cauca, durante temporadas los pobladores percibieron gran cantidad de personas con diferentes discapacidades (Franco 2000). En consideración con lo anterior, el vínculo de diferentes investigadores de la facultad de medicina de la Universidad del Valle y de la Universidad de California UC DAVIS, para el estudio de las causas de la presencia del considerable número de personas con retardo mental (42 de 1.024 habitantes, correspondientes a 27 hombres, 15 mujeres y 10 menores de 14 años), permitió encontrar bajo la aplicación de la prueba genética Linker CGG FMR1 a la mayoría de la población, gran número de personas con diagnóstico SXF, estimando así una prevalencia de SXF para este corregimiento, de al menos 1 en 38 hombres y 1 en 100 mujeres, excediendo a la de la población general en más de 100 veces (Payan et al. 2001).

Por otro lado, el desconocimiento acerca del trastorno ha sido un problema importante para su diagnóstico, ya que, debido a una serie de características en común compartida con otros trastornos o síndromes genéticos, hace que a menudo el SXF no sea diagnosticado o en otros casos deducir diagnósticos erróneos. De esta manera, predomina la necesidad de realizar pruebas genéticas. Sin embargo, el difícil acceso a ellas tanto por cuestiones económicas como por desplazamiento, lleva a pensar que la prevalencia del síndrome puede ser mucho más alta de lo que se cree en la actualidad (Martín 2015).

En este sentido, por medio de la combinación de algunas variables medidas en el cuestionario "Genética Médica Poblacional" que se aplicó en paralelo con la prueba genética en el estudio de Saldarriaga et al. (2014); más la inclusión de los resultados obtenidos en la aplicación del test WHODAS 2.0, y los índices de discapacidad construidos en el estudio previo, los cuales pretenden dar medición a la variable latente discapacidad, los investigadores buscan analizar como por medio de algunas manifestaciones clínicas observables o de información familiar de antecedentes con el trastorno, entre otras (medidas en el cuestionario) es posible estimar la probabilidad de presencia o ausencia de los tres posibles clases de trastorno (zona gris, premutación y mutación completa).

3. Objetivos

3.1. Objetivo General

Ajustar un modelo que permita estimar la probabilidad de presentar un diagnostico normal versus las alteraciones mutación completa, premutación y zona gris, a partir de covariables asociadas a antecedentes de la enfermedad, fenotipo sugestivo y la variable latente grado de discapacidad, aproximada mediante tres métodos diferentes.

3.2. Objetivo Específicos

- Ajustar un modelo binario considerando las variables antecedentes de la enfermedad, fenotipo sugestivo y seleccionar aquellas significativas.
- Ajustar tres diferentes modelos en donde cada uno incluye la variable latente con cada método construido y las variables de interés.
- Establecer el modelo que mejor estime la probabilidad de que un individuo presente alteración mutación completa, premutación o zona gris.

4. Justificación

En Colombia según el estudio realizado por Payan et al. (2001) el pequeño corregimiento de Bolívar, Valle, es la región que hasta el momento evidencia más casos de personas con SXF en el mundo (1 en 38 hombres y 1 en 100 mujeres), cifras obtenidas gracias a los avances tecnológicos y al trabajo investigativo llevado a cabo por los profesionales médicos. Además, su estudio permitió principalmente conocer la causa de la gran prevalencia de retardo mental. Por otro lado, a través de la administración de la encuesta "Genética Médica Poblacional" se obtuvo información acerca de las características demográficas, estado de salud y antecedentes familiares asociados al trastorno de los habitantes del corregimiento. Sin embargo, fuera de la investigación, el costo asociado a estas pruebas es un factor limitante y es ahí donde se cree que la prevalencia del SXF puede ser mayor.

Conforme a Saldarriaga et al. (2014) las pruebas para realizar el diagnóstico de SXF deben ser solicitadas a pacientes con déficit intelectual y/o autismo y además debe haber presente alguna de las siguientes características: orejas grandes y aladas, cara alargada, entre otras; historia familiar de déficit intelectual, autismo, entre otras. No obstante, en diferentes casos estas manifestaciones clínicas no son evidentes y más aún en el caso de las mujeres. Personas con ataxia (torpeza o pérdida de coordinación), síntomas neurológicos o falla ovárica prematura son candidatos a para realizar pruebas moleculares para SXF. Sin embargo, debe considerarse la carencia de centros de salud especializados para diagnóstico de estas enfermedades o eventos. En consecuencia, éste trabajo busca a partir de a estimación de la probabilidad de presencia de los tres posibles trastornos, proporcionar un filtro confiable que permita identificar los individuos que por sus características se les debe aplicar la prueba genética y a quienes no.

5. Antecedentes

5.1. Antecedente contextual

En su estudio Cedillo et al. (2014) propuso estimular al clínico buscar en los pacientes con retraso mental, trastorno por déficit de atención (TDA) e hiperactividad para encontrar el SXF, ya que los diagnósticos incorrectos complican el tratamiento. Este trabajo presentó la revisión literaria del SXF y un caso clínico de una paciente femenina de 9 años sin antecedentes patológicos, es decir, sin historial previa de enfermedades. Esta paciente presentó características clínicas asociadas al SXF como cara alargada, orejas desproporcionadas, frente amplia, entre otras. Además, respecto a sus características cognitivas se mostró pasiva y sin signos de desesperación. Finalmente, en su estudio resaltó la importancia del diagnóstico oportuno de SXF, debido a que muchos de los pacientes que se presentaron a la consulta odontológica tienen TDA e hiperactividad, atribuido a los cambios sociales y culturales que estamos viviendo y no al SXF. Es decir, muchos de los niños mal diagnosticados pasan como niños normales el cual dificulta como se mencionó anteriormente los tratamientos necesarios.

Kidd et al. (2014) presentó en su revisión literaria, informacion actualizada sobre los problemas médicos asociados con SXF. Además, presentó una comparación de la prevalencia registrada en los estudios revisados con los datos recopilados de 9 clínicas especializadas en frágil X de Estados Unidos entre 2005 y 2011 que atienden niños y adultos con el trastorno. En la revisión literaria se encontró que el trastorno puede conducir a presentar alteraciones cardíacas, gastrointestinales, neurológicos, oculares, del oído, la nariz, la garganta, hasta problemas para dormir. La comparación reveló en la prevalencia de otitis media recurrente, una alteración del oído, valores similares, aproximadamente 50 % y convulsiones de 10 % a 16 %, más prevalente en pacientes masculinos. Por el contrario, en las demás alteraciones se revelaron amplias diferencias.

5.2. Antecedente estadístico

El síndrome metabólico (MS) es una agrupación de al menos tres de las siguientes condiciones médicas: aumento de la presión arterial, alto nivel de azúcar en la sangre, obesidad, perímetro abdominal y triglicéridos séricos altos, que aumenta el riesgo de desarrollar enfermedades cardiovasculares y diabetes. Gunderson et al. (2009) estudia 1451 mujeres que nunca han dado a luz, con edades entre 18-30 años y sin diagnóstico de MS. En su estudio estimó los riesgos relativos de MS, por medio del modelo log-log complementario entre los grupos de nacimientos ajustados por raza, edad y otras variables de seguimiento. En su ajuste el modelo evidenció que la maternidad se asoció directamente el desarrollo futuro de MS, independientemente de la obesidad previa y el aumento de peso relacionado con el embarazo. Además, la raza no evidenció tener asociación en los grupos de nacimiento e incidencia de MS

Torréns et al. (2009) estudió 3302 mujeres entre 42-52 años, con condiciones tales como tener útero y al menos un ovario, no hacer uso de estrógenos y al menos un periodo menstrual en los tres meses previos del estudio. Además, se obtuvo información demográfica, antecedentes de hipertensión o diabetes, entre otros. Donde a partir del modelo con enlace log-log complementario se evaluó la asociación entre el MS y los niveles hormonales. Con los resultados no rechazo la hipótesis inicialmente planteada que el cambio en la relación testosterona/estradiol (T/E) durante la transición menopáusica estaría asociado con el síndrome metabólico y el cambio a lo largo del tiempo en el exceso relativo de andrógenos se asociaron con un mayor riesgo de desarrollar el síndrome metabólico durante la transición menopáusica.

6. Marco teórico

6.1. Marco teórico clínico

6.1.1. Síndrome X Frágil

También conocido como síndrome de Martin-Bell, es un trastorno causada por la alteración de un gen específico. Normalmente, el gen produce una proteína necesaria para el desarrollo cerebral. Pero un defecto en este gen hace que una persona produzca poco o nada de dicha proteína, esto resulta en el síndrome de X frágil ocasionando discapacidad mental, problemas de inteligencia que van desde problemas de aprendizaje hasta la discapacidad intelectual grave además de problemas emocionales y sociales (Lorente 2017).

Causas

El SXF según Lorente (2017), es causado por un cambio en la secuencia normal del ADN, en el gen FMR1 (Fragile X Mental Retardation-1). El cambio más frecuente consiste en el alargamiento de una pequeña parte de su secuencia formada por la repetición de las bases nitrogenadas Citosina-Guanina-Guanina (CGG).

La proteína producida FMRP (Fragil X syndrome Mental Retardation Protein) por el gen FMR1, que necesita el cerebro para su correcto funcionamiento, ante dichos defectos del gen FMR1 hace que el cuerpo produzca muy poco de esta proteína o nada en absoluto.

Sobre la base de repeticiones GCG, se pueden formar cuatro categorías alélicas de FMR1, *Normal* (5 - 44 repeticiones); *Intermedio o Zona gris* (45 - 54 repeticiones); *Premutación* (55 - 200 repeticiones); *Mutación completa* (Más de 200 repeticiones).

Diagnóstico

Biggers (2016) señala que para el diagnóstico del SFX, los niños pueden mostrar signos de retraso en el desarrollo u otros síntomas externos, como un gran perímetro cefálico o sutiles diferencias en los rasgos faciales y antecedentes familiares, en este sentido se recomienda someterse a pruebas usando un análisis de sangre de ADN llamado la prueba de ADN FMR1. Esta prueba busca cambios en el gen FMR1 que están asociados con SFX, la cual puede optar por hacer pruebas adicionales para determinar la gravedad de la condición.

Tratamiento

A pesar de que SXF no tiene cura, es posible tratar algunos síntomas con terapia educativa, de la conducta o física y medicinas para la ansiedad y/o depresión. Cuando la enfermedad se trata a partir de la niñez temprana los niños pueden llegar desenvolverse mejor.

6.1.2. Discapacidad

Cabrera (2012) establece que las personas con discapacidad son aquellas con deficiencias físicas, mentales, intelectuales o sensoriales que impiden su participación plena y efectiva en la sociedad.
Además, señala las siguientes cuatro clasificaciones de la discapacidad:

- **Discapacidad física o motriz**
  Incluye personas que presentan en forma permanente debilidad muscular, perdida o ausencia de alguna parte de su cuerpo, alteraciones articulares o presencia de movimientos involuntarios.

- **Sensorial**
  Comprende a las personas con deficiencias visuales, auditivas y a quienes presentan problemas en la comunicación y el lenguaje (ceguera y sordera).

- **Intelectual**
  En este caso se presenta disminución de las funciones mentales superiores (inteligencia, lenguaje y aprendizaje, entre otros), así como de las funciones motoras. Aquí se encuentran las personas que presentan discapacidades para aprender y para realizar diferentes actividades de la vida diaria.

- **Discapacidad psicosocial**
  La discapacidad psicosocial puede convertirse en una condición de vida, que puede ser temporal o permanente. En este se encuentre, por ejemplo, la depresión, esquizofrenia, condición de vida.

6.1.3. Test de discapacidad WHODAS 2.0

El WHODAS 2.0, es un instrumento de evaluación genérico, desarrollado para evaluar las dificultades debidas a problemas de salud, como enfermedades, problemas mentales o emocionales. Es un cuestionario de autoadministrado que evalúa las limitaciones de la actividad y las restricciones de participación (es decir, discapacidad), este test se desarrolló para evaluar seis tareas diferentes de la vida adulta, llamados dominios: 1) comprensión y comunicación; 2) cuidado personal; 3) Movilidad; 4) Relaciones interpersonales; 5) Funciones del trabajo y del hogar; y 6) roles comunitarios y cívicos

(participación) (Organización Mundial de la Salud 2010).

El test se encuentra desarrollado en dos versiones, 12 y 36 preguntas. Estas versiones evalúan el grado de discapacidad de una persona a partir de los cambios en el funcionamiento y sus niveles de dificultad y/o limitación para el desempeño de actividades de la vida diaria, así como, las consecuentes restricciones de participación (Zepeda & Vásquez 2012).
Conforme a la Organización Mundial de la Salud (2010) el test WHODAS 2.O, tras diferentes estudios de confiabilidad y validez demostró presentar sólido respaldo teórico para proporcionar un perfil y una medición total del funcionamiento y la discapacidad, confiable y aplicable interculturalmente en todas las poblaciones adultas.

Puntaje del WHODAS 2.0

Organización Mundial de la Salud (2010) señala las siguientes dos opciones para el cálculo de los puntajes totales para la versión corta (12 preguntas) o versión completa (36 preguntas).

- **Puntaje simple:** Se suman los puntajes asignados a cada una de las preguntas, recodificados previamente como "ninguno" (0), "leve" (1) "moderado" (2) "severo" (3) y "extremo" (4). Es llamado así, ya que los puntajes son sumados sin agrupar las categorías de respuesta o dominios y, por lo tanto, no hay ponderación de tales elementos.

- **Puntaje complejo:** De manera análoga, los puntajes son recodificados y en este caso, permite un análisis más detallado utilizando toda la información de las categorías por separado y luego usa un algoritmo para determinar la puntuación del resumen al ponderar diferencialmente los ítems y los niveles de gravedad. La puntuación tiene tres pasos:
  Paso 1: Suma de puntajes de elementos recodificados dentro de cada dominio.
  Paso 2: Suma de los seis puntajes de dominio.
  Paso 3: Conversión de la calificación resumida en una métrica que va de 0 a 100 (donde 0 = sin discapacidad, 100 = discapacidad completa).

6.2. Marco teórico estadístico

6.2.1. Variable latente

Define Everett (2013) como variable latente, aquella variable que no es observable directamente, y que además no existe un método operativo para la medición directa.
Es decir, una variable latente es un constructo supuesto, en nuestro caso la discapacidad, que sólo puede ser medida mediante la recopilación de varias cantidades medibles llamadas variables manifiestas, por ejemplo, el test de discapacidad.

El análisis de factores pertenece a una familia de métodos multivariados que involucran lo que se denomina variables latentes. Galbraith et al. (2008) menciona que el análisis factorial es el método más antiguo y ampliamente utilizado para inferir las variables latentes, su objetivo es investigar la dependencia de un conjunto de variables manifiestas en un pequeño número de variables latentes (factores) e identificarlos (Everett 2013).

Es así como según Avella et al. (2010) las variables latentes también pueden ser medidas a través de índices sintéticos, construidos a partir de los métodos estadísticos multivariados, como los son AF, ACP, entre otros.

6.2.2. Indicadores e índices

Los indicadores permiten resumir una gran cantidad de datos para dar información de diferentes situaciones presentes en las poblaciones. Estos se convierten en una herramienta cuantitativa simplificada para dar explicación de un atributo a evaluar (Escobar 2006).

Por otra parte, indica que al construir indicadores se debe distinguir entre indicadores simples e indicadores sintéticos o índices. Los indicadores simples están constituidos por la combinación de dos o más datos y los indicadores sintéticos son resultantes de la combinación de los indicadores simples mediante una función matemática que los sintetiza. En cuestión al poder explicativo de los índices, éste no logra dar explicación de todos los factores que pueden describir una variable latente; sin embargo, son una aproximación a ella (Zarzosa 1996).

6.2.3. Curva ROC

Se definen previamente los siguientes conceptos para dar comprensión de la curva ROC.

Un Test de referencia o Gold Standard es considerada como una prueba que proporciona la mejor alternativa diagnóstica para estudiar una determinada enfermedad o evento de interés (Salech et al. 2008). Sin embargo, en diferentes ocasiones los test de referencia son de difícil acceso y por lo tanto se desarrollan otros test diagnósticos de más fácil acceso que permitan sustituir los test de referencia procurando que estos proveen resultados confiables tal como lo hace el test de referencia (Pita Fernández & Pértegas Díaz 2003).

La calidad de una prueba diagnóstica se evidencia en la precisión para clasificar individuos como enfermos o no enfermos, sin embargo, dicha prueba puede presentar fallas en su clasificación y dividirá la muestra de individuos en cuatro subgrupos como se presenta en la siguiente tabla de contingencia:

Tabla 6-1.: Resultado de una prueba y su clasificación respecto al verdadero estado de la enfermedad

		Gold Standard (GS)	
		Enfermo	Sano
Prueba	Enfermo	Verdadero Positivo (VP)	Falso Positivo (FP)
	Sano	Falso Negativo (FN)	Verdadero Negativo (VN)

Una manera convencional para evaluar la calidad de un test según Hajian-Tilaki (2013) es usar la sensibilidad y especificidad como medidas de precisión.

- **Sensibilidad:** Es la probabilidad que tiene una prueba diagnóstica para clasificar a un individuo enfermo, cuando realmente está enfermo:

$$S = Pr(Prueba = enfermo|GS = enfermo) = \frac{VP}{VP + FN} \tag{6-1}$$

  Denominada también, como *fracción de verdaderos positivos* (FVP).

  Al aplicar, la definición "la probabilidad de un suceso es igual a 1 menos la probabilidad de su complementario" se tiene que:

$$1 - S = 1 - \frac{VP}{VP + FN} = \frac{FN}{VP + FN} \tag{6-2}$$

  conocido como la *fracción de falsos negativos* (FFN).

- **Especificidad:** Es la probabilidad que tiene una prueba diagnóstica para clasificar a un individuo sano, cuando realmente está sano.

$$E = Pr(Prueba = sano|GS = sano) = \frac{VN}{VN + FP} \tag{6-3}$$

  Denominada como *fracción de verdaderos negativos* (FVN).

  Luego, al aplicar el complemento se tiene:

$$1 - E = 1 - \frac{VN}{VN + FP} = \frac{FP}{VN + FP} \tag{6-4}$$

  conocido como la *fracción de falsos positivos* (FFP).

Sin embargo, estas medidas de precisión varían en función del punto de corte elegido entre la población sana y la enferma. Por lo tanto, una forma global de conocer la calidad de una prueba diagnóstica es a través de la **curva ROC** (Burgueño et al. 1995).

La curva ROC es una herramienta gráfica bidimensional que representa en el eje vertical los valores de sensibilidad o fracción de verdaderos positivos (FVP) y en el eje horizontal los valores de 1-especificidad o fracción de falsos positivos (FFP), para cada posible punto de corte en la escala de resultados del test de estudio. Puesto que en ambos ejes se tiene probabilidades, la curva ROC estará contenida en el cuadrado unitario $[0, 1] \times [0, 1]$ (Valle Benavides 2017).

El análisis de la curva ROC se suele utilizar para evaluar el poder discriminatorio de un test diagnóstico continuo para clasificar a los individuos en dos grupos diferentes, generalmente enfermos y sanos.

Métodos de cálculo de la curva ROC

Métodos no paramétricos:
Caracterizados por no hacer suposiciones sobre la distribución de los resultados de la prueba diagnóstica. Dentro de los métodos no paramétricos para la construcción de las curvas ROC se encuentra:

- **Método empírico**
  Este método hace uso de las funciones de distribución empíricas de los individuos que están enfermos y los que no. Consiste en representar todos los pares (1-especificidad, sensibilidad) para todos los posibles puntos de corte para la construcción de la curva ROC (Valle Benavides 2017).
  La representación de la curva ROC obtenida por este método, provee una forma escalonada, indicando los cambios producidos en la sensibilidad o especificidad, para cada variación mínima del valor de corte (López de Ullibarri & Pita 1998).

Métodos paramétricos:
Consiste en determinar una distribución para la variable respuesta de la prueba. Señala López de Ullibarri & Pita (1998) que la distribución binormal, que supone la normalidad de las variables tanto en la población sana como en la enferma, es usada con más frecuencia. Sin embargo, también pueden usarse otras distribuciones, similares a la distribución normal como la logística, la exponencial negativa, entre otras.
Una vez elegido la distribución, deben estimarse sus parámetros por un método estadísticamente adecuado, como el método de máxima verosimilitud. Donde finalmente con los parámetros la distribución asumida proporciona una representación de la curva ROC y bajo este enfoque se obtiene una curva ROC suave (Valle Benavides 2017).

Medidas de exactitud

Construida la curva ROC, las medidas de exactitud permiten valorar si la prueba diagnóstica discrimina de manera correcta los individuos. Algunas medidas de exactitud son:

- Exactitud o AC (por su nombre en inglés)

  Es la probabilidad que tiene una prueba para discriminar correctamente, y en términos de las frecuencias observadas se define como:

  $$\hat{AC} = \frac{VP + VN}{VP + VN + FP + FN} \tag{6-5}$$

  Equivalente a decir la proporción de resultados acertados por la prueba.

- Área bajo la curva

  El área bajo la curva (AUC), por sus siglas en inglés, es la probabilidad de clasificar correctamente individuos como enfermos o sanos. El AUC toma un valor entre 0 y 1, donde la exactitud máxima de un test corresponde a un valor de AUC igual a 1 y mínima de 0.5 (Pérez Fernández 2015). Una alta exactitud se evidencia gráficamente en la curva ROC al observar un desplazamiento "hacia arriba y a la izquierda" (López de Ullibarri & Pita 1998).

En situaciones donde se tiene un solo conjunto de datos y una sola curva ROC, el investigador generalmente solo tiene la necesidad de evaluar que tan buena es la prueba o el test para discriminar.

Sin embargo, señala Hanley & McNeil (1983) que cuando se tienen dos pruebas diagnósticas se necesitan criterios estadísticos más formales para juzgar si las diferencias observadas en la exactitud tienen más probabilidades de ser al azar o por realmente tener mayor capacidad discriminante. Lo que lleva a plantear la siguiente hipótesis:

Comparación de pruebas

$$\begin{cases} H_0: & AUC_1 = AUC_2 \\ H_1: & AUC_2 \neq AUC_2 \end{cases} \tag{6-6}$$

Hanley & McNeil (1983) proponen el siguiente estadístico para el contraste de hipótesis

$$z = \frac{\hat{AUC}_1 - \hat{AUC}_2}{SD(\hat{AUC}_1 - \hat{AUC}_2)} \tag{6-7}$$

Donde SD es la desviación estándar y su cálculo viene dado por la siguiente ecuación

$$SD(\hat{AUC}_1 - \hat{AUC}_2) = \sqrt{\hat{var}(\hat{AUC}_1) + \hat{var}(\hat{AUC}_2) - 2 \cdot r \cdot SD(\hat{AUC}_1) \cdot SD(\hat{AUC}_2)} \quad (6\text{-}8)$$

Donde r es el coeficiente de correlación entre ambas áreas. Bajo la hipótesis nula, el estadístico z sigue una distribución normal $N(0,1)$. Entonces para un nivel de confianza α se rechaza la aleatoriedad del test si $|z| > z_{\alpha/2}$
Finalmente para determinar el punto de corte que proporciona la sensibilidad y especificidad más alta, es aquel que presenta el mayor **índice de Youden *(J)***, definido como:

$$J = sensibilidad + especificidad - 1$$

Gráficamente, éste corresponde al punto de la curva ROC más cercano al ángulo superior-izquierdo del gráfico (punto X=0, Y=1), es decir, más cercano al punto del gráfico cuya sensibilidad = 100 % y especificidad = 100 % (Cerda & Cifuentes 2012).

6.2.4. El método de remuestreo: Bootstrap

Fue una técnica propuesta por Efron (1992) y la denominación de remuetreo se debe a que el método se basa en la extracción de un gran número de muestras con repetición de los datos que constituyen la muestra original. En este sentido, la técnica bootstrap es un plan de remuestreo, la cual permite estudiar el comportamiento de los estimadores y el error estadístico, ya sea en cuanto al sesgo, error estándar o tasa de error en una predicción (Solanas & Olivera 1992).

El término puede referirse a bootstrap no paramétrico o bootstrap paramétrico. Los métodos paramétricos implican que el muestreo se realice a partir de una distribución de probabilidad completamente especificada. En el caso no paramétrico, la distribución no se especifica (Gil Abreu 2014). La forma en la cual se desarrolla el método *bootstrap* no paramétrico usado en este trabajo viene dado por los siguientes pasos:

1. El investigador selecciona un estadístico de interés, sea la media, mediana, correlación, tasas, tasa de error, etc.

2. Sea $X = (x_1, x_2, ..., x_n)$ la muestra de datos con unidades $x_1, x_2, ..., x_n$, se asigna a cada unidad la probabilidad de suceso $1/n$, es decir, cada unidad tiene la misma probabilidad de ser seleccionada aleatoriamente.

3. Se define una muestra bootstrap como una muestra aleatoria de tamaño $n^* = n$ de la siguiente manera:
$$X^* = (x_1^*, x_2^*, ..., x_n^*)$$
La notación estrella $(^*)$ indica que (X^*) es una submuestra de X.

4. Lo anterior se repite $\boldsymbol{B}$ ocasiones, y a esto se le denomina $\boldsymbol{B}$ muestras bootstrap.

5. Para cada muestra bootstrap $X^{*1}, X^{*2}, ..., X^{*B}$ se obtiene el estadístico de interés, denotado como

$$\hat{\theta}^{*1}, \hat{\theta}^{*2}, ..., \hat{\theta}^{*B}$$

El conjunto de estimaciones θ^{*b} para $b = 1, 2, ..., B$, constituye la estimación *bootstrap* de la distribución muestral del estadístico de interés.

La estimación bootstrap del error estándar o desviación estándar de la distribución muestral de un estadístico, el cual es el error en la estimación del estadístico de interés a partir de las muestras, se obtiene mediante la expresión:

$$\widehat{SD_{boot}} = \left\{ \sum_{b=1}^{B} \frac{(\hat{\theta}^{*b} - \hat{\theta}_{boot})^2}{B-1} \right\}^{1/2}$$

Donde

$$\hat{\theta}_{boot} = \sum_{b=1}^{B} \frac{\hat{\theta}^{*b}}{B}$$

6.2.5. Modelo lineal

Faraway (2014) define un modelo lineal como un esquema de relación lineal entre una variable aleatoria Y (dependiente o exógena) a explicar y un conjunto de variables no aleatorias X (X_1, X_2,...,X_p) (endógenas o explicativas). Cuya estructura general de análisis con variables dependiente continua es:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_p X_{pi} + \varepsilon_i \tag{6-9}$$

Para $i = 1, 2, ..., n$. Donde (X_{1i}, X_{2i}, ..., X_{pi}, Y_i) son características asociadas al i-ésimo individuo, β_0, β_1,...,β_p son parámetros desconocidos del modelo, la cual miden la influencia que las variables explicativas tienen sobre la variable dependiente y ε es el termino aleatorio de error. Respecto a los supuestos de la variable dependiente, las observaciones Y_i provienen de una distribución normal con media μ_i y varianza σ^2, respectivamente y los siguientes para los errores residuales:

- Especificación del modelo:

  $E(\varepsilon_i) = 0$ para $i = 1, 2, ..., n$

  Finalmente se modela la relación lineal de la media de la variable de respuesta con el conjunto de variables explicativas, esto es:

$$\mu_i = E(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_p X_{pi} \tag{6-10}$$

- Homogeneidad de varianza de los errores, supone que la variable dependiente (Y) presenta una distribución con igual varianza en todo el rango de valores de la variable independiente (X).

  $V\big(\varepsilon_i\big) = \sigma^2$ para $i = 1, 2, ..., n$

  o equivalente a decir, $V\big(Y_i\big) = \sigma^2$ para $i = 1, 2, ..., n$

- No correlación de los errores

  $COV\big(\varepsilon_i, \varepsilon_{i'}\big) = 0 \ \forall i \neq i'$.

- Distribución normal de los errores

  $\varepsilon_i \sim Normal\big(0, \sigma^2\big)$

El éxito en el ajuste de un modelo de regresión y la validez de los hallazgos y conclusiones obtenidas, dependen del cumplimiento de estos supuestos (Gutiérrez 2003).

6.2.6. Modelo lineal generalizado

Los modelos lineales generalizados (MLG) son una generalización de los modelos de regresión lineal estándar antes visto, empleados por falta de linealidad y homogeneidad de la varianza, ya que ciertos tipos de variables de respuesta sufren al menos uno de los supuestos de los modelos normales antes mencionados (Cayuela 2009).

Agresti (2015) menciona tres componentes que especifican los MLG:

- El *componente aleatorio* identifica la variable respuesta (Y) y su distribución de probabilidad o función de masa para una distribución en la familia exponencial; además las observaciones de Y $(y_1, y_2, ..., y_n)$ en la distribución supuesta se tratan como independientes.

- El *componente sistemático*, también llamado predictor lineal, especifica las variables explicativas la cual se incorporan linealmente para relacionarse con el nivel de Y_i.
  $\eta_i = \sum_{j=1}^{p} \beta_j X_{ji}$

- *Función de enlace*, una función $g(\cdot)$ del valor esperado de Y $(\mu_i = E(Y_i))$, monótona, diferenciable que relaciona la esperanza de la variable dependiente con el predictor lineal.
  $\eta_i = g(\mu_i)$

Finalmente se obtiene un modelo de la forma:

$$\eta_i = g\big[E(y_i)\big] = \sum_{j=1}^{p} \beta_j X_{ji} \tag{6-11}$$

Una de las ventajas que señala Agresti (2015) al restringir los GLM a las distribuciones pertenecientes a la familia exponencial como lo son la distribución normal, binomial, multinomial y la distribución poisson, se obtienen expresiones más generales para las ecuaciones de verosimilitud del modelo, distribuciones asintóticas de los estimadores para los parámetros del modelo y un algoritmo para ajustar los modelos.

Familia Exponencial

Se definió anteriormente el componente aleatorio de un GLM, la cual consiste en que la variable respuesta Y presenta observaciones $(y_1, y_2, ..., y_n)$ independientes, con función de densidad o de masa que pueden escribirse de la forma:

$$f(y_i; \theta_i, \phi) = exp\left\{ \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\} \tag{6-12}$$

Donde θ es llamado parámetro natural y ϕ es denominado como parámetro de dispersión. El parámetro natural relaciona la media y como su nombre lo indica el parámetro de dispersión relaciona la varianza. La forma algebraica de la ecuación 6-12 se le llama familia de dispersión exponencial (Agresti 2015)

6.2.7. Estimación

El método máxima verosimilitud, permite obtener las estimaciones de los parámetros de los modelos lineales generalizados y su procedimiento descrito por Hardin et al. (2007), viene dado de la siguiente manera:

Como primera instancia, se define la función de densidad conjunta de las observaciones de la muestra y partiendo del supuesto que son independientes, la distribución conjunta de las observaciones dado los parámetros θ y ϕ viene dada por:

$$f(y_1, y_2, ..., y_n; \theta, \phi) = \prod_{i=1}^{n} exp\left\{ \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\} \tag{6-13}$$

Siendo las observaciones de la muestra conocidas, pero el parámetro desconocido la función que expresa los parámetros θ y ϕ como cantidades aleatorias, permitiendo realizar inferencias acerca de su valor a partir de un conjunto de observaciones y_i, es llamada función de verosimilitud y es escrita como:

$$L(\theta, \phi; y_1, y_2, ..., y_n) = \prod_{i=1}^{n} exp\left\{ \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\} \tag{6-14}$$

Por medio de una transformación a la función de verosimilitud L, establecida de forma multiplicativa, se logra obtener una función en una escala aditiva, conocida como la función

de log verosimilitud, la cual provee los mismos resultados que aquellos que maximizan la función de verosimilitud.

$$\mathcal{L} = \sum_{i=1}^{n} \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + \sum_{i=1}^{n} c(y_i, \phi) \tag{6-15}$$

Para obtener las ecuaciones de máxima verosimilitud (EMV) de β, se debe maximizar la ecuación 6-15. Y para ello, se plantea inicialmente la derivada respecto al parámetro de interés β que se obtendrá aplicando la regla de la cadena para una muestra de n observaciones.

$$\frac{\partial \mathcal{L}}{\partial \beta_j} = \sum_{i=1}^{n} \left(\frac{\partial \mathcal{L}_i}{\partial \theta_i}\right)\left(\frac{\partial \theta_i}{\partial \mu_i}\right)\left(\frac{\partial \mu_i}{\partial \eta_i}\right)\left(\frac{\partial \eta_i}{\partial \beta_j}\right) \tag{6-16}$$

Las derivadas parciales individuales son obtenidas partiendo de las siguientes propiedades

$$E\left(\frac{\partial \mathcal{L}}{\partial \theta_i}\right) = 0 \tag{6-17}$$

$$E\left(\frac{\partial^2 \mathcal{L}}{\partial \theta_i^2}\right) + E\left(\frac{\partial \mathcal{L}}{\partial \theta_i}\right)^2 = 0 \tag{6-18}$$

Donde, de la ecuación 6-15 se obtiene la primera y segunda derivada, dada por:

$$\frac{\partial \mathcal{L}}{\partial \theta_i} = \frac{y_i - b^{'}(\theta_i)}{a(\phi)} \tag{6-19}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \theta_i^2} = \frac{-b^{''}(\theta_i)}{a(\phi)} \tag{6-20}$$

Luego, de la primera derivada (ecuación 6-19), se obtiene la expresión de $E(y_i)$

$$\begin{aligned} E\left(\frac{\partial \mathcal{L}}{\partial \theta_i}\right) &= 0 \\ &= E\left(\frac{y_i - b^{'}(\theta_i)}{a(\phi)}\right) = 0 \\ b^{'}(\theta_i) &= E(y_i) = \mu_i \end{aligned} \tag{6-21}$$

De la segunda derivada (ecuación 6-20), se obtiene la expresión de $V(y_i)$

$$\begin{aligned} E\left(\frac{-b^{''}(\theta_i)}{a(\phi)}\right) + E\left(\left(\frac{y_i - b^{'}(\theta_i)}{a(\phi)}\right)^2\right) &= 0 \\ \frac{-b^{''}(\theta_i)}{a(\phi)} + \frac{1}{a(\phi)^2} E\{y_i - \mu_i\}^2 &= 0 \\ \frac{-b^{''}(\theta_i)}{a(\phi)} + \frac{1}{a(\phi)^2} V(y_i) &= 0 \\ V(y_i) &= b^{''}(\theta_i) a(\phi) \end{aligned} \tag{6-22}$$

De $\mu_i = b^{'}(\theta_i)$ luego como $V(y_i) = b^{''}(\theta_i)a(\phi)$ entonces

$$\frac{\partial \mu_i}{\partial \theta_i} = b^{''}(\theta_i) = \frac{V(y_i)}{a(\phi)} \quad \text{luego} \quad \frac{\partial \theta_i}{\partial \mu_i} = \frac{a(\phi)}{V(y_i)} \tag{6-23}$$

Para $\eta_i = g(\mu_i)$

$\frac{\partial \mu_i}{\partial \eta_i}$ depende de la función de enlace $g(\cdot)$

Con $\eta_i = \sum_{j=1}^{p} \beta_j x_{ji}$

$$\frac{\partial \eta_i}{\partial \beta_j} = x_{ji}$$

Por lo tanto, se tiene finalmente para la formulación de la ecuación 6-16,

$$\frac{\partial \mathcal{L}}{\partial \beta_j} = \sum_{i=1}^{n} \left(\frac{y_i - b^{'}(\theta_i)}{a(\phi)} \right) \left(\frac{a(\phi)}{V(y_i)} \right) \left(\frac{\partial \mu_i}{\partial \eta_i} \right) x_{ji}$$

Finalmente, la estimación de máxima verosimilitud para β_j para $j = 1, 2, .., p$ se obtiene resolviendo las ecuaciones:

$$\frac{\partial \mathcal{L}}{\partial \beta_j} = \sum_{i=1}^{n} \frac{(y_i - \mu_i) x_{ji}}{V(y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0 \tag{6-24}$$

Métodos de aproximación

Ya que las ecuaciones de verosimilitud de la ecuación 6-24 son generalmente no lineales en β. La estimación de los parámetros requiere de métodos numéricos, generalmente, la estimación de máxima verosimilitud se lleva a cabo con la puntuación de Fisher (también llamada mínimos cuadrados ponderados iterativos) que es una modificación del algoritmo de Newton-Raphson.

Newton-Raphson es un método iterativo para resolver ecuaciones no lineales, como por ejemplo, determinar el punto en el cual la función toma su máximo. Las estimaciones de máxima verosimilitud son generadas a partir de una secuencia de aproximaciones de la formula

$$\beta^{(t+1)} = \beta^{(t)} - (H^{(t)})^{-1} u^{(t)} \tag{6-25}$$

La cual convergen a la ubicación del máximo cuando la función es adecuada y la aproximación inicial fue buena. Donde $H^{(t)}$ y $u(t)$, son H y u evaluadas en $\beta^{(t)}$, con:

$$H = \left[\frac{\partial^2 \mathcal{L}(\beta)}{\partial \beta_a \beta_b} \right]_{a,b=1,2,...,p}$$

Es la matriz de información observada, llamada matriz de Hesse y

$$u = \left(\frac{\partial \mathcal{L}(\beta)}{\partial \beta_1}, \frac{\partial \mathcal{L}(\beta)}{\partial \beta_2}, ..., \frac{\partial \mathcal{L}(\beta)}{\partial \beta_p} \right) \tag{6-26}$$

Método puntación de Fisher, a diferencia del método de Newton-Raphson utiliza el valor esperado de la matriz de informacion observada, esto es:

$$\mathcal{J} = -E \left[\frac{\partial^2 \mathcal{L}(\beta)}{\partial \beta_a \beta_b} \right]_{a,b=1,2,...,p} \tag{6-27}$$

evaluadas en $\beta^{(t)}$, las estimaciones de los parámetros son generadas a partir de la secuencia de aproximaciones de la formula

$$\beta^{(t+1)} = \beta^{(t)} + (\mathcal{J}^{(t)})^{(-1)} u^{(t)} \tag{6-28}$$

6.2.8. Modelo lineal generalizado para datos binarios

Señala Flórez & Rincón (2012) que algunas variables de respuesta Y son de tipo dicotómica, es decir Y es una variable que toma solo dos posibles valores.
En este caso, el propósito se basa en encontrar la probabilidad que un individuo manifieste éxito dado un conjunto de covariables. Esto es:

$$Y_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_p X_{pi} + \varepsilon_i$$

La distribución de Y se especifica por las probabilidades $P(Y = 1) = \theta$ de éxito y $P(Y = 0) = 1 - \theta$ de fracaso, con esperanza $E(Y) = \theta$.
Lo que implica bajo el supuesto $E(\varepsilon_i) = 0$ que:

$$E(Y_i|X_i) = \theta_i = X_i'\beta$$

Lo anterior denota que la respuesta esperada del modelo es la probabilidad de que la variable dependiente tome el valor de 1. Es por dicha cuestión que un modelo con variable dependiente dicotómica, llamado generalmente modelo lineal de probabilidad presente alguna de las siguientes dificultades:

- No cumplimiento de $0 \leq E(Y_i|X_i) \leq 1$, ya que debe tenerse en cuenta que las variables explicativas pueden ser de naturaleza cualitativa o cuantitativa.
- Los términos de error, solo pueden tomar los valores $\varepsilon_i = 1 - X_i'\beta$ si $y_i = 1$ y $\varepsilon_i = -X_i'\beta$ si $y_i = 0$. Por lo tanto, los errores no se distribuyen según la distribución normal.
- La varianza del error es función de los valores de X, es decir no es constante, puesto que $var(\varepsilon_i) = \theta_i(1 - \theta_i)$. En presencia de heterocedasticidad, los estimadores MCO, además de insesgados no son eficientes; es decir no tienen varianza mínima y por lo tanto no se obtendrán resultados válidos.

De modo a lo anterior, cuando son usadas ciertas variables de respuesta como las de conteo de casos, respuesta binaria, en modelos lineales los supuestos tienden a no cumplirse. Es así como los **modelos lineales generalizados (GLM)** son una alternativa para transformar la variable respuesta, dada la falta de linealidad y homogeneidad de la varianza (Mayoral & Socuéllamos 2001).

Conforme a McCullagh & Nelder (1989), la tabla **6-2** muestra tres principales funciones de enlaces para modelos de respuesta binaria:

Tabla 6-2.: Funciones de enlaces para modelos de respuesta binaria

Nombre del enlace	**Función de enlace** $\eta = g(\theta)$	**Enlace inverso** $\pi = g^{-1}(\eta)$
Logit	$ln\left(\frac{\theta}{1-\theta}\right)$	$\frac{e^{\eta}}{1+e^{\eta}}$
Probit	$\Phi^{-1}(\theta)$	$\Phi(\theta)$
Log-log complementario	$log(-log(1-\theta))$	$1 - exp\{-exp(\eta)\}$

Enlace log log Complementario

El modelo binario con función de enlace log log complementario tiene la forma:

$$\eta_i = log(-log(1-\theta_i)) = \sum_{j=1}^{p} \beta_j X_{ji} \tag{6-29}$$

Donde la expresión del lado izquierdo es la transformación log log al complemento de la probabilidad θ, la cual toma una respuesta restringida en el intervalo $(0, 1)$ y la traslada al intervalo real $(-\infty, +\infty)$. Por tanto $-\infty < \eta_i < \infty$ y $\beta_1, \beta_2, ..., \beta_p$ son los coeficientes desconocidos.
A diferencia de los enlaces Logit y Probit presentados en la tabla **6-2**, la función de enlace log log complementaria (cloglog) no es simétrica alrededor de $\theta = 0{,}5$, es decir, la probabilidad de éxito no es la misma a la probabilidad de fracaso, y además dado que la distribución del valor extremo es asimétrica, los resultados son ligeramente diferentes de los otros dos modelos simétricos (Logit y Probit). Es así como de este modo, el modelo cloglog se basa en la distribución de tipo extremo I, también denominada distribución gumbel (Mauchant et al. 2011).
Kotz & Nadarajah (2000) indica que la *Distribución de gumbel (Distribución de valor extremo generalizado tipo I)* es usada para modelar la distribución del máximo o el mínimo de una muestra, con función de distribución acumulada:

$$F(x) = Pr[X \leq x] = exp\big\{ - exp[(x-\mu)/\sigma]\big\} \tag{6-30}$$

Donde $-\infty < \mu < \infty$ y $\sigma > 0$ son los parámetros de localización y escala respectivamente.

Dada una variable de respuesta binaria, los eventos raros binarios se refiere al número pequeño de éxitos observados de la variable respuesta, y teniendo en cuente que los modelos Logit y Probit son simétricos, en presencia de eventos raros estos muestran inconvenientes ya que la probabilidad del evento raro se subestima (Calabrese et al. 2011)
Luego para obtener los valores ajustados en el intervalo $(0, 1)$

$$\theta_i = 1 - exp[-exp(\sum_{j=1}^{p} \beta_j X_{ji})] \tag{6-31}$$

La función inversa del log-log complementario presentada en la ecuación 6-31, es uno menos la función de distribución acumulativa de la variable aleatoria de Gumbel.

6.2.9. Análisis del ajuste

Desviación

Determinar que tan bien el modelo lineal generalizado con enlace cloglog describe los datos observados, se le denomina juzgar la adecuación del modelo y un método para tal efecto es comparar el logaritmo de la verosimilitud del modelo con enlace *cloglog* ajustado con el logaritmo de la verosimilitud de un modelo saturado, la cual es un modelo que se ajusta perfectamente a los datos de la muestra con n parámetros. La desviación de acuerdo con Montgomery et al. (2001) viene expresada como:

$$\lambda(\beta) = 2\big[\mathcal{L}(\text{modelo saturado}) - \mathcal{L}(\hat{\beta})\big] \tag{6-32}$$

Bajo el supuesto que el modelo cloglog es la función correcta y el tamaño n de la muestra es grande, entonces:

$$\lambda(\beta) \sim \chi^2_{n-p}$$

Es decir, chi-cuadrado con $n - p$ grados de libertad.

La ecuación 6-33 presenta la regla de decisión, para determinar cuando el modelo es correcto

$$\begin{aligned} &\text{si } \lambda(\beta) \leq \chi^2_{\alpha;n-p} \text{ el modelo ajustado es adecuado} \\ &\text{si } \lambda(\beta) \geq \chi^2_{\alpha;n-p} \text{ el modelo ajustado no es adecuado} \end{aligned} \tag{6-33}$$

6.2.10. AIC

Es una medida de la calidad relativa de un modelo estadístico, para un conjunto observaciones, de tal forma el AIC es un medio para seleccionar un modelo. El objetivo

es encontrar el modelo que tenga la más baja perdida de información. En este sentido, los valores más bajo del criterio de un modelo preferible.

$$AIC = 2p - 2\mathcal{L}(M_k) \tag{6-34}$$

Donde p es el número de parámetros en el modelo estadístico y $\mathcal{L}(M_k)$ es la log verosimilitud para el modelo estimado k. La medida de criterio es adecuada para comparar los GLM del mismo enlace con diferentes listas de covariables.

7. Metodología

En el presente estudio se propone por medio de los modelos lineales generalizados identificar variables que permitan predecir el estado de la alteración genética mutación completa, premutación y zona gris, tomando como evento complementario para cada una "NO" presenta alteración genética, es decir, resultado "Normal" de la prueba genética. Para alcanzar este objetivo se trabajó con los datos del estudio Saldarriaga et al. (2014) llevado a cabo por diferentes profesionales de la salud, en un corregimiento al occidente del Valle del Cauca. De donde se obtuvo información del estado de la alteración zona gris (ZG), premutación (PM), mutación completa (MC) o estado normal (Normal), a través de la aplicación de la prueba genética Linker CGG FMR1. Anexo a la prueba genética, la administración del cuestionario "Genética Médica Poblacional", proporcionó información de característica demográficas, antecedentes familiares con SXF, estado de salud y discapacidad.

Dicho lo anterior, este trabajo complementa los resultados obtenidos en un estudio previo desarrollado por Lopez & Diaz (2017) en el cual, con base en los resultados de la prueba genética Linker CGG FMR1 y el cuestionario "Genética Médica Poblacional", se construyó un índice de discapacidad mediante dos aproximaciones metodológicas, el proceso de jerarquía analítica (PJA) y por métodos estadísticos multivariados a través de los resultados de la pregunta 20 del cuestionario. Los autores usaron el test WHODAS 2.0 como instrumento para validación de criterio.

7.1. Descripción del cuestionario "Genética Médica Poblacional" y covariables a seleccionar

El cuestionario "Genética Médica Poblacional", está compuesto por 20 preguntas, la cual suministran información del sexo, características demográficas básicas, socioeconómicas, antecedentes familiares con SXF, estado de salud y discapacidad; valoradas sobre preguntas puntuales, aplicadas a personas diagnosticadas con los tres tipos de alteración y sin la alteración (normal).

Del cuestionario "Genética Médica Poblacional" y teniendo en cuenta que el SXF en sus formas son de tipo genético hereditaria, la cual pueden manifestarse generalmente en los niveles de discapacidad intelectual y en otras clínicas, se seleccionan la pregunta 18:

"¿Fenotipo sugestivo de Síndrome X Frágil?" y la pregunta 19: "¿Probables o confirmados casos de Frágil X en su familia?" como variables explicativas para el ajuste del modelo lineal generalizado.

Por otro lado, los índices de discapacidad construidos en el estudio previo, usando el método proceso de jerarquía analítica (PJA) y métodos estadísticos multivariados, análisis de componentes principales, a partir de los resultados de la pregunta 20 compuesto por los siguientes 9 *ítems*: moverse o caminar, usar brazos y/o manos, oír aún con aparatos especiales, hablar o comunicarse, ver a pesar de usar lentes o gafas, entender o aprender, relacionarse con los demás por problemas mentales, emocionales o de nervios, desplazarse en trechos cortos por problemas del corazón o respiratorios y autocuidado. También serán usados como variables explicativas cada uno en diferentes modelos, así como el índice de discapacidad obtenido por el test WHODAS 2.0, con el objetivo de determinar el más significativo para dar mejor pronóstico de los trastornos.

7.2. Descripción de los datos, "Test WHODAS 2.0"

El test WHODAS 2.0 (WDS2.0), proporciona una escala de evaluación del grado de discapacidad validado por OMS. Es un test estandarizado, que permite dar una mayor dimensión de la población con discapacidad, además permite identificar los diferentes grados de impacto (leve, moderado y severo) junto con las principales dificultades de funcionamiento (Zepeda & Vásquez 2012).

Dicho lo anterior, este estudio obtiene los puntajes de discapacidad de las personas a partir de la aplicación del test WDS2.0 versión 12 preguntas, útil para evaluaciones breves del funcionamiento general en encuestas, de manera análoga a la versión completa, éste permite calcular puntuaciones de funcionamiento general y éste explica el 81 % de la varianza de la versión de 36 elementos (Organización Mundial de la Salud 2010).

Finalmente, los puntajes totales fueron obtenidos por medio del cálculo complejo. Para ello se tuvo en cuenta la composición de los dominios para la versión de 12 preguntas, D1: pregunta 1 y 2; D2: pregunta 3 y 4; D3: pregunta 5 y 6; D4: pregunta 7 y 8; D5: pregunta 9 y 10 y D6: pregunta 11 y 12.

7.3. Descripción de los índices y sus puntos de corte

Con el propósito de diferenciar los índices, éstos fueron nombrados de la siguiente manera:

- índice-1: Índice de discapacidad obtenido del test WDS2.0
- índice-2: Índice de discapacidad construido por medio del método estadístico multivariado ACP.
- índice-3: Índice de discapacidad construido por medio del proceso de jerarquía analítica (PJA).

7.3.1. índice-1

Organización Mundial de la Salud (2001) menciona los puntos de corte del test WDS2.0 para clasificar a las personas en sus diferentes niveles de discapacidad.

Tabla 7-1.: Puntos de corte del índice-1

	No hay problema	Problema leve	Problema moderado	Problema grave	Problema completo
WDS2.0	$0 - 4\,\%$	$5\,\% - 24\,\%$	$25\,\% - 49\,\%$	$50\,\% - 95\,\%$	$96\,\% - 100\,\%$

Para determinar los puntos de corte de los índices 2 y 3, Lopez & Diaz (2017) usaron las curvas ROC como método para ello. Las siguientes subsecciones presentan los resultados que obtuvieron:

7.3.2. índice-2

De acuerdo con el índice de Youden obtenido con un valor de 0,43, el mejor punto de corte para clasificar personas con discapacidad grave a moderada es 0,62, con sensibilidad de 0,48 y especificidad 0,95. Luego, el mejor punto de corte para clasificar a las personas con discapacidad moderara a leve fue 0,03 con un índice de Youden de 0,43, sensibilidad 0,73 y especificidad 0,73.

Tabla 7-2.: Puntos de corte del índice-2

	No discapacidad	Discapacidad leve	Discapacidad moderada	Discapacidad grave
ACP	0	$< 0{,}03$	$< 0{,}62$	$\geq 0{,}62$

7.3.3. índice-3

Con un valor de 0,99 el índice de Youden, sugirió que el mejor punto de corte para diferenciar entre los niveles de discapacidad grave a moderada fuera 0,33, proporcionando una sensibilidad de 0,99 y una especificidad de 1. Por otro lado, para clasificar personas con niveles de discapacidad moderada a leve el punto de corte fue 0,11 con un índice de Youden de 0,71 el cual presentó sensibilidad de 0,97 y especificidad de 0,74.

Tabla 7-3.: Puntos de corte del índice-3

	No discapacidad	Discapacidad leve	Discapacidad moderada	Discapacidad grave
PJA	0	$< 0{,}11$	$< 0{,}33$	$\geq 0{,}33$

7.4. Descripción de la muestra

La aplicación del test WDS2.0 se llevó a cabo mediante un diseño muestral propuesto por Lopez & Diaz (2017) en su estudio previo, usando como marco muestral los individuos con resultados de la prueba genética y el cuestionario aplicado conjuntamente a gran parte de la población. Así pues, las personas que seleccionaron al menos un ítem de la pregunta 20 y presentaron resultados zona gris, premutación, mutación completa o normal por la prueba genética, fueron censadas para la administración del test WDS2.0. Así mismo, las personas que no seleccionaron ningún ítem de la pregunta 20, pero que aun así tienen diagnóstico mutación completa, premutación o zona gris por la prueba genética.

Para aquellas personas que no seleccionaron ningún ítem de la pregunta 20 y presentaron diagnóstico normal por la prueba genética, el test WDS2.0 fue aplicado a una muestra aleatoria de ellas obtenida por medio de una enumeración previa a cada individuo con las características mencionadas y la generación de números aleatorios por medio del programa estadístico R (R Core Team 2017). En el caso en que dicha persona muestreada ya no viviera en el corregimiento o hubiera fallecido, se propuso una lista adicional de personas seleccionadas aleatoriamente para reemplazarlos.

La administración del test se realizó bajo un único encuestador debidamente capacitado, en cuanto a la importancia de los datos, el objetivo del test y la forma de ser diligenciado.

Finalmente, el archivo de datos consta de la información de los resultados de la prueba genética, del cuestionario “Genética Médica Poblacional” y del test WDS2.0 para 284 personas.

7.5. Modelo estadístico

Experimento estadístico

Aplicar el cuestionario "Genética poblacional", la prueba genética y el test WHODAS 2.0 y observar las respuestas y el resultado.

Eventos y variable aleatoria

Se considera como variable respuesta de interés el tipo de trastorno y para su modelación se definen los siguientes eventos de interés, las variables aleatorias y finalmente su distribución asociada.

Los posibles resultados de la variable respuesta de interés, viene dado por los siguientes eventos:

$$A_1 = \{\text{individuo con MC}\} \qquad A_1^C = \{\text{individuo normal}\} \tag{7-1}$$
$$A_2 = \{\text{individuo con PM}\} \qquad A_2^C = \{\text{individuo normal}\} \tag{7-2}$$
$$A_3 = \{\text{individuo con ZG}\} \qquad A_3^C = \{\text{individuo normal}\} \tag{7-3}$$

Luego, se define Y_k como la variable aleatoria que identifica la presencia o ausencia del evento A_k, para $k = \{1, 2, 3\}$.

$$Y_k = \begin{cases} 1 & \text{si ocurre} A_k \\ 0 & \text{si ocurre} A_k^C \end{cases} \tag{7-4}$$

Con distribución asociada, $Y_k \sim Bernoulli(\theta_k)$, donde θ_k es el parámetro que representa la probabilidad de presentar el trastorno A_k, para $k = \{1, 2, 3\}$.

7.6. Descripción de las variables explicativas

Éste estudio propone modelar en forma separada la probabilidad de ocurrencia de los tres posibles eventos de interés, MC, PM y ZG (éxitos) versus el evento estado Normal (fracaso) para cada caso. Teniendo en cuenta lo anterior, esta sección presenta las variables independientes asociadas a cada evento éxito, para su respectiva modelación.

Debe tenerse en cuenta que la información de las personas diagnosticadas como Normal, es la misma información de fracaso, que se tiene en cuenta en los tres diferentes éxitos. Luego de manera general para $k = \{1, 2, 3\}$ con 1= MC y Normal, 2= PM y Normal, 3= ZG y Normal, se tiene:

Tabla 7-4.: Variables independientes, significado y naturaleza

Variable	Significado	Tipo	Naturaleza
$X_1^{(k)}$	Es el vector de respuesta a la pregunta 18 hecha a las personas con $k = \{1, 2, 3\}$.	Cualitativa	Binaria 1: Positivo 0: Negativo
$X_2^{(k)}$	Es el vector de respuesta a la pregunta 19 hecha a las personas con $k = \{1, 2, 3\}$.	Cualitativa	Binaria 1: Positivo 0: Negativo o No sabe
índice-$1^{(k)}$	Es el vector de resultados del índice de discapacidad obtenido de la aplicación del test WDS2.0 para las personas con $k = \{1, 2, 3\}$.	Cuantitativa	Continua
índice-$2^{(k)}$	Es el vector de resultados del índice de discapacidad obtenido por el método ACP para las personas con $k = \{1, 2, 3\}$.	Cuantitativa	Continua
índice-$3^{(k)}$	Es el vector de resultados del índice de discapacidad obtenido por el método PJA para las personas con $k = \{1, 2, 3\}$.	Cuantitativa	Continua

7.7. Descripción de los modelos lineales generalizados a comparar

Cuando las variables dependientes binarias estudiadas presentan un número muy pequeño de éxitos, como sucede en este trabajo, es decir presencia de asimetría, se recomienda usar el modelo binario con enlace log-log complementario (*cloglog*) para predecir la probabilidad de ocurrencia de los eventos en función de las covariables definidas anteriormente.

Se ajustaron cuatro modelos diferentes para predecir la probabilidad de ocurrencia de cada uno de los eventos, y por medio del menor estadístico AIC se seleccionó el modelo que mejor predijo cada probabilidad.

Los modelos a ajustar tienen la forma:

$$\eta^{(k,l)} = g(\theta_k) = ln(-ln(1 - \theta_k)) \tag{7-5}$$

Donde los super índices (k, l) hacen identificar al modelo l que se ajustó para el tipo de trastorno k, de modo que $l = \{1, 2, 3, 4\}$ e $k = \{1, 2, 3\}$.

Tabla 7-5.: MLG's para modelar la ocurrencia del trastorno MC

θ_1
$\eta_i^{(1,1)} = \beta_0^{(1)} + \beta_1^{(1)} X_{1i}^{(1)} + \beta_2^{(1)} X_{2i}^{(1)}$
$\eta_i^{(1,2)} = \beta_0^{(1)} + \beta_1^{(1)}$índice-$1_i^{(1)}$
$\eta_i^{(1,3)} = \beta_0^{(1)} + \beta_1^{(1)}$índice-$2_i^{(1)}$
$\eta_i^{(1,4)} = \beta_0^{(1)} + \beta_1^{(1)}$índice-$3_i^{(1)}$

Para $i = \{1, 2, 3, ..., 230\}$.

Tabla 7-6.: MLG's para modelar la ocurrencia del trastorno PM

θ_2
$\eta_i^{(2,1)} = \beta_0^{(2)} + \beta_1^{(2)} X_{1i}^{(2)} + \beta_2^{(2)} X_{2i}^{(2)}$
$\eta_i^{(2,2)} = \beta_0^{(2)} + \beta_1^{(2)}$índice-$1_i^{(2)}$
$\eta_i^{(2,3)} = \beta_0^{(2)} + \beta_1^{(2)}$índice-$2_i^{(2)}$
$\eta_i^{(2,4)} = \beta_0^{(2)} + \beta_1^{(2)}$índice-$3_i^{(2)}$

Para $i = \{1, 2, 3, ..., 241\}$.

Tabla 7-7.: MLG's para modelar la ocurrencia del trastorno ZG

θ_3
$\eta_i^{(3,1)} = \beta_0^{(3)} + \beta_1^{(3)} X_{1i}^{(3)} + \beta_2^{(3)} X_{2i}^{(3)}$
$\eta_i^{(3,2)} = \beta_0^{(3)} + \beta_1^{(3)}$índice-$1_i^{(3)}$
$\eta_i^{(3,3)} = \beta_0^{(3)} + \beta_1^{(3)}$índice-$2_i^{(3)}$
$\eta_i^{(3,4)} = \beta_0^{(3)} + \beta_1^{(3)}$índice-$3_i^{(3)}$

Para $i = \{1, 2, 3, ..., 239\}$.

Si alguna o ambas covariables ($X_1^{(k)}$ o $X_2^{(k)}$) resultan significativas en el ajuste de los modelos cloglog $\eta^{(k,1)}$, para la explicación de la probabilidad éxito $k = \{1, 2, 3\}$, entonces dicha variable o ambas entrarán en los modelos cloglog $\eta^{(k,2)}$, $\eta^{(k,3)}$ y $\eta^{(k,4)}$ los cuales fueron ajustados con la información contenida en los índices (1,2 y 3) respectivamente, como covariables.

7.8. Evaluación de la capacidad predictiva de los modelos

Consecutivo a la evaluación de los modelos propuestos y haber seleccionado aquellos estadísticamente significativos para dar predicción para cada tipo de trastorno, se procedió a valorar que tan bueno es el modelo discriminando a los individuos con algún trastorno y sin trastorno. Para ello, se construyó para cada modelo seleccionado una curva ROC no paramétrica por el método empírico, donde a partir del índice Youden encontrar su correspondiente punto de corte, el cual proporciona las mejores medidas de exactitud.

Entonces, sea $\hat{\theta}_i^{(k,l)}$ el valor predicho del individuo i, con el modelo (k,l). Sea la variable aleatoria de prueba $Z^{(k,l)}$ con distribución Bernoulli, de la que se estudiará su eficiencia discriminante, con resultados:

$$Z^{(k,l)} = \begin{cases} 1 & \text{si } \hat{\theta}_i^{(k,l)} \geq \theta_0^{(k,l)} \\ 0 & \text{si } \hat{\theta}_i^{(k,l)} < \theta_0^{(k,l)} \end{cases} \tag{7-6}$$

Entonces la variable aleatoria $Z^{(k,l)}$ tomará resultados positivo ($Z^{(k,l)} = 1$) o negativo ($Z^{(k,l)} = 0$) en función del valor $\theta_0^{(k,l)}$ denominado punto de corte o valor umbral de los valores predichos del modelo l ajustado para predecir el trastorno k.

A partir de la variable aleatoria Z definida anteriormente, se evaluó la clasificación de los modelos seleccionados. El proceso implicó comparar los valores predichos del modelo versus los valores observados para cada observación, y una forma para ello es a partir de la construcción de una tabla de clasificación que cruza la respuesta binaria (presencia o ausencia) de cada tipo de trastorno resultante de la prueba genética con la clasificación binaria predicha. La cual tiene la forma presentada en la tabla **7-8**.

Tabla 7-8.: Número de individuos para el trastorno k correcta e incorrectamente clasificados

	Prueba genética (Gold Standard)	
	$Y_k = 1$	$Y_k = 0$
$Z^{(k,l)} = 1$	Verdadero Positivo (VP)	Falso Positivo (FP)
$Z^{(k,l)} = 0$	Falso Negativo (FN)	Verdadero Negativo (VN)

Donde:

- (VP): es el número de positivos para cada trastorno bien clasificados por el modelo.
- (FP): es el número de positivos para cada trastorno mal clasificados por el modelo.
- (FN): es el número de negativos para cada trastorno mal clasificados por el modelo.
- (VN): es el número de negativos para cada trastorno bien clasificados por el modelo.

7.9. Evaluación de la capacidad clasificatoria de los modelos

La evaluación de la propiedad clasificatoria de los modelos tiene como objetivo determinar si los modelos para cada trastorno pueden clasificar de forma precisa los individuos con y sin el evento a partir de nueva información. Para evaluar la capacidad clasificatoria de los modelos se procedió a desarrollar la metodología de bootstrap que permite obtener estimadores no sesgados del desempeño en el futuro del modelo sin sacrificar el tamaño de la muestra. Los pasos son descritos a continuación y se aplicaron para cada trastorno.

1. Asignar a cada vector de observaciones $\left[Y_i, X_{1i}, X_{2i}, \text{índice-1}_i, \text{índice-2}_i, \text{índice-3}_i\right]$ teniendo en cuenta el i-ésimo individuo de la muestra, se asigna la probabilidad de suceso $1/n$, es decir, cada individuo tiene la misma probabilidad de ser seleccionada aleatoriamente.

2. Una vez seleccionada una muestra aleatoria con sustitución de tamaño igual a la muestra original ($n^* = n$), se obtienen las medidas de interés: exactitud (AC), (FVP), (FVN), (FFP) y (FFN).

3. Generar $B = 1000$ muestras bootstrap independientes de tamaño $n^* = n$ empleando el muestreo con sustitución. Esto es:

$$n^{*1}, n^{*2}, ..., n^{*1000}$$

4. Para cada muestra b-ésima se obtienen las medidas descritas en el paso 2.

$$\widehat{AC}^{*1}, \widehat{AC}^{*2}, ..., \widehat{AC}^{*1000}$$
$$\widehat{FVP}^{*1}, \widehat{FVP}^{*2}, ..., \widehat{FVP}^{*1000}$$
$$\widehat{FVN}^{*1}, \widehat{FVN}^{*2}, ..., \widehat{FVN}^{*1000}$$
$$\widehat{FFN}^{*1}, \widehat{FFN}^{*2}, ..., \widehat{FFN}^{*1000}$$
$$\widehat{FFP}^{*1}, \widehat{FFP}^{*2}, ..., \widehat{FFP}^{*1000}$$

Donde

$$\widehat{AC}^{*b} = \frac{VP^{*b} + VN^{*b}}{VP^{*b} + VN^{*b} + FP^{*b} + FN^{*b}}$$

$$\widehat{FVP}^{*b} = \frac{VP^{*b}}{VP^{*b} + FN^{*b}} \qquad \widehat{FVN}^{*b} = \frac{VN^{*b}}{VN^{*b} + FP^{*b}}$$

$$\widehat{FFN}^{*b} = \frac{FN^{*b}}{VP^{*b} + FN^{*b}} \qquad \widehat{FFP}^{*b} = \frac{FP^{*b}}{VN^{*b} + FP^{*b}}$$

para $b = 1, 2, ..., B$

5. Estimar la medida de exactitud y fracciones bootstrap de la siguiente manera:

$$\widehat{FVP_{boot}} = \sum_{b=1}^{B=1000} \frac{FVP^{*b}}{1000} \qquad \widehat{FFP_{boot}} = \sum_{b=1}^{B=1000} \frac{FFP^{*b}}{1000}$$

$$\widehat{FFN_{boot}} = \sum_{b=1}^{B=1000} \frac{FFN^{*b}}{1000} \qquad \widehat{FVN_{boot}} = \sum_{b=1}^{B=1000} \frac{FVN^{*b}}{1000}$$

$$\widehat{AC_{boot}} = \sum_{b=1}^{B=1000} \frac{AC^{*b}}{1000}$$

La estimación bootstrap de la desviación estándar, es decir, el error en la estimación de la medida de exactitud y fracciones a partir de las muestras, se obtiene mediante la expresión:

$$\widehat{SD(AC)}_{boot} = \left\{ \sum_{b=1}^{B=1000} \frac{(\widehat{AC}^{*b} - \widehat{AC_{boot}})^2}{1000-1} \right\}^{1/2}$$

$$\widehat{SD(FVP)}_{boot} = \left\{ \sum_{b=1}^{B=1000} \frac{(\widehat{FVP}^{*b} - \widehat{FVP_{boot}})^2}{1000-1} \right\}^{1/2}$$

$$\widehat{SD(FVN)}_{boot} = \left\{ \sum_{b=1}^{B=1000} \frac{(\widehat{FVN}^{*b} - \widehat{FVN_{boot}})^2}{1000-1} \right\}^{1/2}$$

$$\widehat{SD(FFN)}_{boot} = \left\{ \sum_{b=1}^{B=1000} \frac{(\widehat{FFN}^{*b} - \widehat{FFN_{boot}})^2}{1000-1} \right\}^{1/2}$$

$$\widehat{SD(FFP)}_{boot} = \left\{ \sum_{b=1}^{B=1000} \frac{(\widehat{FFP}^{*b} - \widehat{FFP_{boot}})^2}{1000-1} \right\}^{1/2}$$

Como la FFN es el complemento de FVP entonces $\sqrt{var(FFN)} = \sqrt{var(FVP)}$. De manera análoga, como FFP es el el complemento de la FVN entonces $\sqrt{var(FFP)} = \sqrt{var(FVN)}$.

8. Resultados

En este capítulo se presentan los resultados del análisis descriptivo de las variables de respuesta de interés, las variables explicativas y el cruce entre ellas. Posteriormente se presentan los resultados de los modelos mediante el modelo cloglog, las validaciones, comparaciones y finalmente el análisis de la capacidad predictiva de los modelos seleccionados.

8.1. Análisis exploratorio de los datos

Este estudio cuenta con un total de 284 personas con información de la prueba genética, el cuestionario "Genética Medica Poblacional", los puntajes del WDS2.0 y los índices construidos por (Lopez & Diaz 2017).

El análisis descriptivo de las variables de interés para evaluar su comportamiento, muestra esencialmente probabilidades bajas de éxitos, ya que de la muestra solo el 5,99 % de las personas presentaron mutación completa, el 9,86 % premutación y el 9,15 % zona gris a comparación con 75, 00 % de las personas diagnosticadas sin ningún trastorno.

Tabla 8-1.: Distribución de la muestra de acuerdo a la prueba genética

Tipo de trastorno	Frecuencia	Porcentaje
Mutacion Completa (MC)	17	5,99 %
Premutación (PM)	28	9,86 %
Zona Gris (ZG)	26	9,15 %
Normal	213	75,00 %
TOTAL	284	100 %

Respecto a las variables explicativas, la tabla **8-2** evidencia que solo el 8,80 % de las veces el encuestador evidenció fenotipo sugestivo o características visibles al síndrome X frágil de la persona a la que encuestó.
Además, ya que gran parte de los individuos de la muestra evidenció no presentar ningún trastorno, es razonable que el 89,79 % de las veces el encuestador no haya evidenciado características sugestivas al SXF, y el 1,41 % de las veces no se pudo diferenciar.

Tabla 8-2.: Distribución de la muestra de acuerdo a la Pregunta 18

Opciones de respuesta	Frecuencia	Porcentaje
Si	25	8,80 %
No	255	89,79 %
No sabe	4	1,41 %
Total	284	100 %

Los resultados obtenidos sobre la pregunta 19 evidencia que 23,24 % de los individuos de la muestra manifestaron tener conocimiento de probables o confirmados casos de SXF en su familia, 75,00 % de los individuos expresaron una respuesta negativa a la pregunta y 1,76 % no tenían suficiente conocimiento para responder (ver tabla **8-3**).

Tabla 8-3.: Distribución de la muestra de acuerdo a la Pregunta 19

Opciones de respuesta	Frecuencia	Porcentaje
Si	66	23,24 %
No	213	75,00 %
No sabe	5	1,76 %
Total	284	100 %

Los resultados obtenidos por el índice-1, mostrados en la tabla **8-4**, evidencia que acorde al bajo porcentaje de personas con los trastornos MC, PM y ZG también se presentan bajos porcentaje de personas con problemas de discapacidad grave (7,39 %), moderado (7,75 %) y leve (8,10 %). Además, el porcentaje de personas sin ningún problema (76,76 %) es aproximadamente igual al número de personas diagnosticadas por la prueba genética como Normal (75,00 %). Lo que supone pensar que el índice-1 esta revelando los problemas de discapacidad presentes en las personas con MC, PM y ZG, como también los no problemas de discapacidad en las personas sin ningún trastorno.

Tabla 8-4.: Distribución de la muestra de acuerdo al índice-1 (WDS2.0)

Estado índice-1 (WDS2.0)	Frecuencia	Porcentaje
Problema Grave	21	7,39 %
Problema Moderado	22	7,75 %
Problema Leve	23	8,10 %
No Hay Problema	218	76,76 %
Total	284	100 %

Los resultados obtenidos del estudio previo donde se propone dar otra medición de la variable latente discapacidad, muestran que la distribución de los niveles de discapacidad hallados con

el índice-2 (Tabla **8-5**), presentan algunas diferencias en cuanto a la distribución de los niveles de discapacidad hallados con el índice-1, ya que, mientras con éste índice se encuentran mas personas con problema leve, el índice-2 muestra menos personas con problema leve (1,76 %) y más personas con problema moderado (11,97 %). No obstante, se observó similitud en la cantidad de personas valoradas sin ningún problema por ambos índices.

Tabla 8-5.: Distribución de la muestra de acuerdo al índice-2 (ACP)

índice-2 (ACP)	Frecuencia	Porcentaje
Problema Grave	16	5,63 %
Problema Moderado	34	11,97 %
Problema Leve	5	1,76 %
No Hay Problema	229	80,63 %
Total	284	100 %

Contrario al índice-2, los problemas de discapacidad hallados con el índice-3 muestran mayor cantidad de personas con problema leve (9,86 %) (ver tabla **8-6**), resultado más acorde con los resultados obtenidos por el test WDS2.0. Sin embargo, con el índice-3 se encuentran menos personas con problema moderado que el hallado con el test WDS2.0 y el índice-2. De igual forma, el índice-3 evidencia más personas sin ningún problema (79, 58 %).

Tabla 8-6.: Distribución de la muestra de acuerdo al índice-3 (PJA)

Estado índice-3 (PJA)	Frecuencia	Porcentaje
Problema Grave	24	8,45 %
Problema Moderado	6	2,11 %
Problema Leve	28	9,86 %
No Hay Problema	226	79,58 %
Total	284	100 %

A continuación, se presentan los resultados del análisis bivariado de las variables explicativas con la variable estado de la persona, es decir, las siguientes tablas de contingencia buscan investigar alguna relación de las variables independientes y la variable dependiente.

La tabla de contingencia **8-7** muestra el porcentaje de veces que el encuestador asoció, no asoció o no supo asociar fenotipo sugestivo del SXF y el individuo presentaba MC, PM, ZG o Normal. Observe como el porcentaje de las veces que el encuestador contestó “Si” y el individuo presentaba MC (4,93 %) fue mayor al porcentaje de veces que el encuestador contestó “Si” y el individuo presentaba PM (3,87 %). Más aún, ninguna de las veces el encuestador asocio fenotipo sugestivo al SXF cuando este se encontró en su forma ZG y cuando los individuos no presentaron ningún trastorno. En este sentido, los resultados

muestran la dificultad de valorar a una persona con ZG por medio de sus características físicas, ya que en estos estados las manifestaciones son menos visibles.

Luego, en los casos cuando el encuestador responde "No", muy pocas veces el individuo presentó MC (1,06 %), ya que fue posible que sus manifestaciones físicas no fueran visibles. Como se mencionó anteriormente, cuando el tipo de trastorno es más leve sus manifestaciones físicas no se asocian claramente, fue así como el porcentaje de las veces que el encuestador negó la pregunta 18 fue mayor cuando el individuo presentaba PM (5, 99 %), ZG (8,80 %) y por ende Normal (73,94 %).

Tabla 8-7.: Distribución de la muestra de acuerdo a la pregunta 18 y tipo de trastorno

Pregunta 18	Tipo de trastorno				
Respuesta	MC	PM	ZG	NORMAL	Total porcentual
Si	4,93 %	3,87 %	0,00 %	0,00 %	8,80 %
No	1,06 %	5,99 %	8,80 %	73,94 %	89,79 %
No sabe	0,00 %	0,00 %	0,35 %	1,06 %	1,41 %
Total porcentual	5,99 %	9,86 %	9,15 %	75,00 %	100,00 %

Finalmente, en el proceso de encuesta el encuestador pudo encontrarse con algunas variables confusoras que no le permitieron definir bien si el individuo presentaba o no fenotipo sugestivo al SXF, esto se evidencia en el porcentaje de veces que el encuestador respondió "No sabe" a la pregunta 18 y el individuo se encontraba en ZG (0,35 %) y sin ningún trastorno (1,06 %).

Los resultados obtenidos en la formulación de la pregunta 19 y el estado del trastorno (ver tabla **8-8**), muestra mayor porcentaje de personas con MC y tener en la familia casos confirmados o sospechados de SXF (5,63 %) que aquellas personas con MC y negar tener casos en la familia de SXF (0,35 %). Es decir, en esta tabla se ve expresada la relación hereditaria de la enfermedad.

Tabla 8-8.: Distribución de la muestra de acuerdo a la pregunta 19 y tipo de trastorno

Pregunta 19	Tipo de trastorno				
Respuesta	MC	PM	ZG	NORMAL	Total porcentual
Si	5,63 %	5,99 %	0,70 %	10,92 %	23,24 %
No	0,35 %	3,87 %	8,45 %	62,32 %	75,00 %
No sabe	0,00 %	0,00 %	0,00 %	1,76 %	1,76 %
Total porcentual	5,99 %	9,86 %	9,15 %	75,00 %	100,00 %

Fueron muy pocos los casos donde los individuos con MC, PM señalaron no tener casos sospechosos o confirmados de SXF en su familia (0,35 % y 3,87 %) respectivamente.

Contrario a las personas con trastorno ZG, ya que estos presentaron mayor porcentaje (8,45 %).
Acorde a la definición del trastorno, todos los individuos con MC, PM y ZG debieron tener antecedentes y por ende haber contestado "Si" a la pregunta 19. Sin embargo, no se rechaza la suposición que las personas que desarrollaron la encuesta y presentaban algún trastorno no tenían suficiente información de esta pregunta y en efecto expresaron una respuesta negativa.

El análisis resultante de la tabla de contingencia de los índices y los tipos de trastorno, muestran que tanto para el índice-1, índice-2 e índice-3 el porcentaje de personas con trastorno MC y problema de discapacidad grave es mayor (4,93 %), (4,58 %) y (4,93 %), respectivamente, que el porcentaje de personas con MC y presentar problema moderado, leve o no tener problema. (Ver tablas **8-9**, **8-10** y **8-11**).

Tabla 8-9.: Distribución de la muestra de acuerdo índice-1 (WDS2.0) y tipo de trastorno

WHODAS 2.0	Tipo de trastorno				
Estado	MC	PM	ZG	NORMAL	Total porcentual
Problema Grave	4,93 %	0,70 %	0,00 %	1,76 %	7,39 %
Problema Moderado	0,70 %	2,82 %	0,35 %	3,87 %	7,75 %
Problema Leve	0,00 %	0,70 %	0,35 %	7,04 %	8,10 %
No Hay Problema	0,35 %	5,63 %	8,45 %	62,32 %	76,76 %
Total porcentual	5,99 %	9,86 %	9,15 %	75,00 %	100,00 %

En la tabla **8-9** se observa que, con el índice-1 el 2,82 % de las personas presentaron PM y problema moderado, lo cual es aproximadamente al mismo resultado que se llega con el índice-2 puesto que el 2,46 % de las personas presentaron PM y problema moderado. A diferencia de los resultados obtenidos con el índice-3, el cual manifiesta que el 2,46 % de las personas presentaron PM y problema grave.

Tabla 8-10.: Distribución de la muestra de acuerdo índice-2 (ACP) y tipo de trastorno

ACP	Tipo de trastorno				
Estado	MC	PM	ZG	NORMAL	Total porcentual
Problema Grave	4,58 %	1,06 %	0,00 %	0,00 %	5,63 %
Problema Moderado	1,06 %	2,46 %	0,70 %	7,75 %	11,97 %
Problema Leve	0,00 %	0,00 %	0,00 %	1,76 %	1,76 %
No Hay Problema	0,35 %	6,34 %	8,45 %	65,49 %	80,63 %
Total porcentual	5,99 %	9,86 %	9,15 %	75,00 %	100,00 %

Se observó finalmente como con los tres índices evidenciar problemas de discapacidad en personas con trastorno ZG es más difícil. Ademas, con los tres índices se observó que

el porcentaje de personas sin ningún trastorno, es decir, "Normal" y con problema de discapacidad grave y moderado fue mayor que observar personas con ZG y algún problema de discapacidad. En este caso es posible que los tres índices se encuentren valorando problemas de discapacidad asociados a otras enfermedades.

Tabla 8-11.: Distribución de la muestra de acuerdo índice-3 (PJA) y tipo de trastorno

PJA	Tipo de trastorno				
Estado	MC	PM	ZG	NORMAL	Total porcentual
Problema Grave	4,93 %	2,46 %	0,00 %	1,06 %	8,45 %
Problema Moderado	0,70 %	0,70 %	0,00 %	0,70 %	2,11 %
Problema Leve	0,00 %	0,35 %	0,70 %	8,80 %	9,86 %
No Hay Problema	0,35 %	6,34 %	8,45 %	64,44 %	79,58 %
Total porcentual	5,99 %	9,86 %	9,15 %	75,00 %	100,00 %

8.2. Ajuste de los modelos con enlace cloglog

Un modelo multinomial puede verse como una generalización del modelo binomial que permite estudiar más de dos posibles resultados. Estos modelos son herramientas de análisis que busca estimar las probabilidades asociadas a cada elección, dadas las características particulares de los individuos o los atributos de las elecciones, resumidas en las covariables. Sin embargo dado los objetivos de este estudio, se ajustaron los modelos con la marginal de los datos multinomial, es decir, usando cada uno de los posibles resultados como binomiales.

Esta sección presenta las estimaciones de los parámetros de los modelos propuestos para la predicción de la presencia o ausencia del SXF en sus formas MC, PM y ZG, la evaluación de la bondad de ajuste de cada uno de ellos, la elección del modelo que será usado y finalmente la valoración de la capacidad predictiva por medio de las curvas ROC.

Los resultados de la estimación por máxima verosimilitud de los parámetros de los modelos cloglog fueron obtenidos a partir del software estadístico R, la cual usa como método de aproximación la puntuación de Fisher, modificación del algoritmo Newton-Raphson.

8.2.1. Mutación Completa como evento éxito

La tabla **8-12** muestra que el error estándar del coeficiente de la variable $X_1^{(1)}$ es grande y además con un nivel de significancia del 0,05 el valor-p (0,9739) asociado al valor z del test de Wald para la prueba de significancia, sugiere no rechazar la hipótesis nula que dicho parámetro sea cero. Es decir que tener o no tener fenotipo sugestivo de SXF no esta

afectando la probabilidad de tener MC.

Sin embargo, el test Wald para el coeficiente de la variable $X_2^{(1)}$ con valor-p = 0,0469 < 0,05, rechaza la hipótesis nula que no hay asociación entre los casos familiares con SXF y el trastorno MC.

Tabla 8-12.: Resultados del modelo $\eta^{(1,1)}$

Componente	$\hat{\beta}$	Error estandar	valor-p	Estadístico	valor	gl
Intercepto	−5,207	1,000	$1{,}92e-07$	D. nulla	121.277	229
$X_1^{(1)}$	5,762	176,148	0,9739	D. residual	27.503	227
$X_2^{(1)}$	2,434	1,225	0,0469			
$AIC = 33{,}503$						
Iteraciones con el algoritmo de Newton Raphson = 17						

De modo que con los resultados obtenidos del modelo $\eta^{(1,1)}$, la variable $X_2^{(1)}$ será incluida en los siguientes modelos donde se encuentran los índices de discapacidad.

Tabla 8-13.: Resultados del modelo $\eta^{(1,2)}$

Componente	$\hat{\beta}$	Error estándar	valor-p	Estadístico	valor	gl
Intercepto	−6,018	1,076	$2{,}25e-08$	D. nulla	121,277	229
$X_2^{(1)}$	3,178	1,060	0,00272	D. residual	41,339	227
índice-$1^{(1)}$	5,954	1,406	$2{,}28e-05$			
$AIC = 47{,}339$						
Iteraciones con el algoritmo de Newton Raphson = 8						

De este modo, todos los coeficientes del modelo $\eta^{(1,2)}$ presentados en la tabla **8-13** fueron significativos para dar predicción del trastorno MC. Además, cada uno de ellos presentó bajo error estándar, es decir se presentó buena precisión en la estimación de los parámetros del modelo. Las estimaciones de los parámetros con signo positivos, muestran que la probabilidad de tener MC aumenta con la información de las covariables $X_2^{(1)}$ e índice-$1^{(1)}$. Por otro lado el modelo propuesto evidenció tener buena bondad de ajuste ya que $\lambda(\hat{\beta}) = 41{,}339$ es menor al valor crítico $\chi^2_{0,05;227} = 193{,}1259$.

En el ajuste del modelo $\eta^{(1,3)}$ con las variables explicativas $X_2^{(1)}$ e índice-$2^{(1)}$ y $\eta^{(1,4)}$ con las variables explicativas $X_2^{(1)}$ e índice-$3^{(1)}$, se evidenció el fenómeno denominado separación cuasi completa. En dicho sentido, la separación completa ocurre cuando una combinación lineal de los predictores es perfectamente predictiva del resultado, conduciendo a grandes estimaciones de los parámetros, tendiendo éstos a ser infinitos con error estándar grandes, en estos casos no existe las estimaciones por máxima verosimilitud (Van der Paal

2014). En contraste la separación casi completa ocurre cuando los datos están cerca de estar perfectamente separados, como ocurre en el caso de los datos con el modelo $\eta^{(1,3)}$ y $\eta^{(1,4)}$.

En consecuencia, cuando una persona tiene mutación completa los problemas de discapacidad son muy evidentes y en este caso se encontró que los índices-2 e índice-3, junto con al información de los casos de SXF en la familia, dieron por sí solos predicción del trastorno mutación completa. En otras palabras, ya que el objetivo de un modelo binario es clasificar los elementos de una muestra en dos grupos excluyentes bajo ciertas características, por ejemplo, enfermos o sanos, y debido a que con altos valores del índice-$2^{(1)}$ e índice-$3^{(1)}$ solo se observaron personas con MC (ver figura **8-1** y **8-2**) y por el contrario con valores bajos solo se encontraron personas sin MC, los modelos no ajustan bien con estas variables ya que evidentemente se encontró separación.

Figura 8-1.: Cuasi separación índice-2

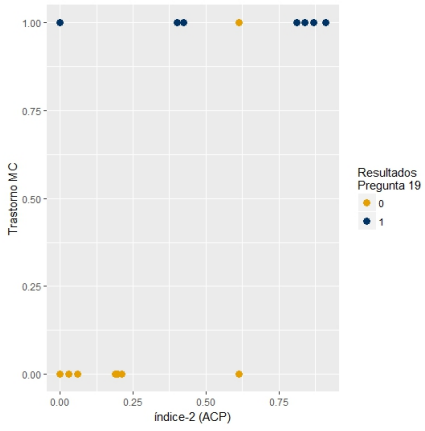


Figura 8-2.: Cuasi separación índice-3

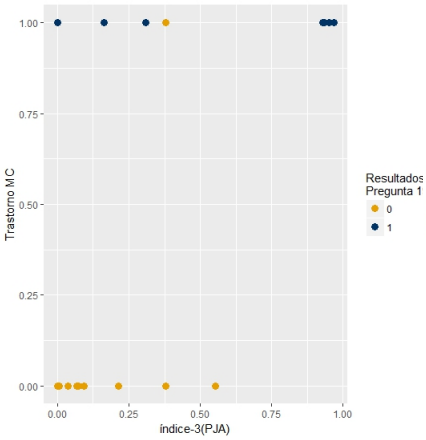
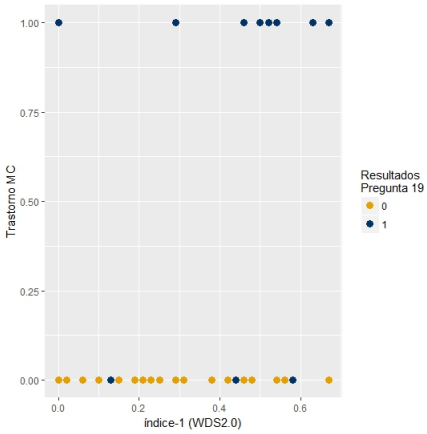


Figura 8-3.: Ausencia de separación índice-1



Contextualmente, el SXF en su forma MC demuestra que no necesita de modelos para ser predicho, puesto que sus manifestaciones físicas como en la discapacidad son bastantes solidas.

Por otro lado, se observó que las covariables incluidas en el modelo $\eta^{(1,2)}$ construido de igual forma para predecir el trastono MC no presentó problemas de separabilidad, ya que como puede verse en la figura **8-3** tanto para valores bajos y altos del índice-$1^{(1)}$ hay personas sin MC, es decir con éste índice se exhibe más incertidumbre. Debido a que el WDS2.0 puede estar evaluando otros problemas de discapacidad no asociados al SXF, se recalca la importancia de la covariable $X_2^{(1)}$ en el modelo para disminuir dicha incertidumbre de asociar o no los problemas de discapacidad al SXF.

8.2.2. Premutación como evento éxito

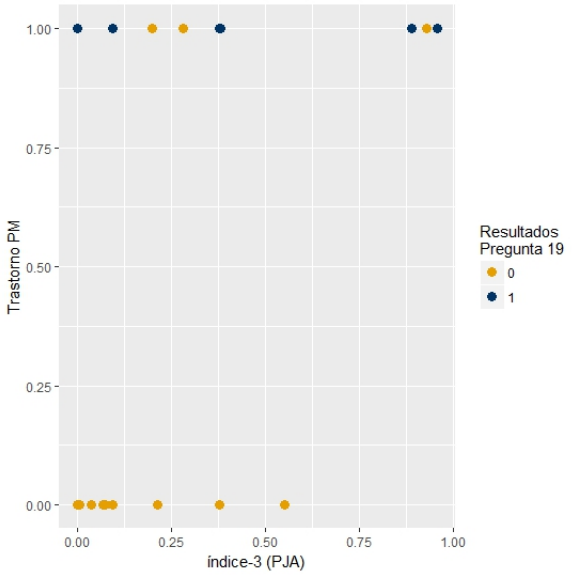
De manera análoga a los resultados anteriores, la tabla **8-14** muestra que la variable $X_1^{(2)}$ en el modelo $\eta^{(2,1)}$ no ayuda a dar predicción del trastorno PM.

Tabla 8-14.: Resultados del modelo $\eta^{(2,1)}$

Componente	β	Error estándar	valor-p	Estadístico	valor	gl
Intercepto	$-2{,}8356$	0,3016	$< 2e-16$	D. nulla	173,16	240
$X_1^{(2)}$	4,6127	79,6842	0,9538	D. residual	117,19	238
$X_2^{(2)}$	1,1036	0,5080	0,0298			
$AIC = 123{,}19$						
Iteraciones con el algoritmo de Newton Raphson = 15						

En el mismo sentido, la covariable $X_2^{(2)}$ significativa para dar predicción del trastorno PM, fue incluida en los modelos $\eta^{(2,2)}$, $\eta^{(2,3)}$ y $\eta^{(2,4)}$. Sin embargo, en el ajuste del modelo $\eta^{(2,4)}$ con covariables índice-3$^{(2)}$ y $X_2^{(2)}$ se presentó lo antes denominado separación cuasi completa (ver figura **8-4**), esto es, a pesar de que los problemas de discapacidad de los individuos con trastorno PM se encontraron más distribuidos en los diferentes niveles, como se vió en las tablas cruzadas de la sección 8.1. La figura muestra que el índice-3$^{(2)}$ obtenido por el método PJA, asigna menor valor solo a las personas sin PM y mayor valor a las personas con PM y a dos individuos sin PM. Por lo tanto, el proceso de estimación de los parámetros de este modelo se elude y se presentan las estimaciones de modelos $\eta^{(2,2)}$, $\eta^{(2,3)}$.

Figura 8-4.: Cuasi separación índice-3



La tabla **8-15** y **8-16** muestran las estimaciones de los parámetros de los modelos $\eta^{(2,2)}$,

$\eta^{(2,3)}$, respectivamente. En ellas, los valores p resultantes del test Wald sugiere que tanto los coeficientes del modelo $\eta^{(2,2)}$ y $\eta^{(2,3)}$ difieren de cero y por ende las variables son significativas en cada modelo para la predicción del trastorno PM (valores p $< 0{,}05$).

Tabla 8-15.: Resultados del modelo $\eta^{(2,2)}$

Componente	β	Error estandar	valor-p	Estadístico	valor	gl
Intercepto	$-2{,}9683$	0,3131	$< 2e-16$	D. nulla	173,16	240
$X_2^{(2)}$	1,8565	0,3958	$2{,}73e-06$	D. residual	142,18	238
índice-$1^{(2)}$	2,0297	0,9219	0,0277			
$AIC = 148{,}18$						
Iteraciones con el algoritmo de Newton Raphson = 6						

En cuanto a los errores estándar de ambos modelos, exhiben que las estimaciones de los coeficientes de ambos modelos fueron realizadas con buena precisión, ya que sus errores estándar fueron bajos.

La desviación resultante evidenció que los tres modelos con enlace cloglog propuestos proporcionan un ajuste adecuado para los datos, ya que $\lambda(\hat{\beta}) = 142{,}18$ para el modelo $\eta^{(2,2)}$ (tabla **8-15**) y $\lambda(\hat{\beta}) = 125{,}00$ para el modelo $\eta^{(2,3)}$ (tabla **8-16**) son inferiores al valor crítico $\chi^2_{0,05;238} = 203{,}286$.

Tabla 8-16.: Resultados del modelo $\eta^{(2,3)}$

Componente	β	Error estandar	valor-p	Estadístico	valor	gl
Intercepto	$-3{,}1404$	0,3361	$< 2e-16$	D. nulla	173,16	240
$X_2^{(2)}$	1,8818	0,4259	$9{,}95e-06$	D. residual	125,00	238
índice-$2^{(2)}$	4,4843	0,9857	$5{,}38e-06$			
$AIC = 131$						
Iteraciones con el algoritmo de Newton Raphson = 7						

8.2.3. Zona Gris como evento éxito

Las covariables índices de discapacidad y $X_1^{(3)}$ y $X_2^{(3)}$ no fueron significativas para dar pronóstico del trastorno ZG, es decir, estas variables no muestran fuerte relación con la presencia del trastorno ZG, como para dar pronóstico. Se señala que los resultados de las tablas **8-9**, **8-10** y **8-11** de la sección 8.1 también exhiben dicha conclusión, puesto que el mayor porcentaje de personas con ZG no evidencian problemas de discapacidad. Por lo tanto, aunque una persona tenga trastorno ZG ninguno de los índices pudo dar determinación de ello.

Elección de los modelos

La elección del mejor modelo se realizó por medio del menor valor AIC, en este sentido para dar predicción del PM el mejor modelo es aquel que se incluye el índice-2 (ACP), esto es, $\eta^{(2,3)}$. Sin embargo, dado lo "inmediato" que es para un investigador la aplicación del test WDS2.0 y su respectiva medición, y además, dada las buenas características del modelo $\eta^{(2,2)}$ con el índice-1, se procede a evaluar la capacidad predictiva y clasificatoria de ambos modelos, aquel que incluye el índice-$2^{(2)}$ (ACP) elegido por el AIC y el modelo $\eta^{(2,2)}$.

8.3. Estudio de la capacidad predictiva y clasificatoria

Las siguientes secciones presentan los resultados de las medidas de exactitud halladas de la construcción de la curva ROC de los valores predichos del modelo $\eta^{(1,2)}$, $\eta^{(2,2)}$ y $\eta^{(2,3)}$. Con éstas medidas se buscó evaluar la capacidad que el modelo $\eta^{(1,2)}$ presentó para clasificar a los individuos con y sin MC y la capacidad que los modelos $\eta^{(2,2)}$ y $\eta^{(2,3)}$ presentaron para clasificar a los individuos con y sin PM.

8.3.1. Estudio de la capacidad predictiva y clasificatoria del modelo $\eta^{(1,2)}$ ajustado para dar predicción del trastorno MC

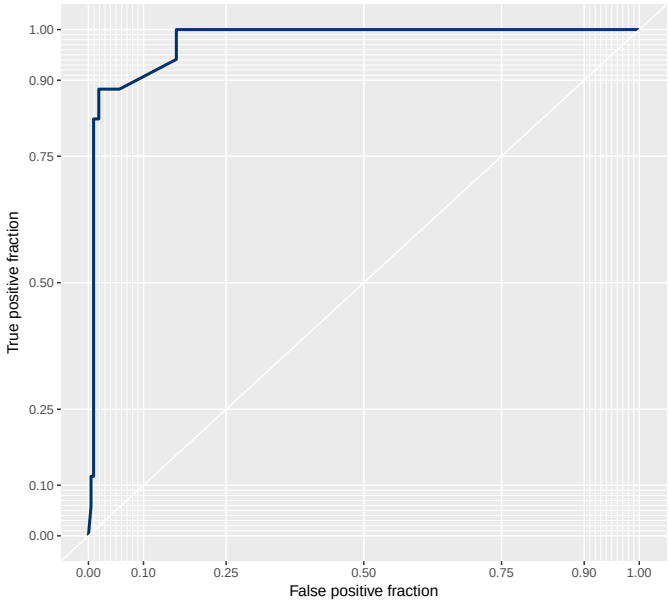
Con las estimaciones de los parámetros del modelo $\eta^{(1,2)}$ presentados en la tabla **8-13**, se tuvieron los valores predichos dados por la siguiente expresión:

$$\hat{\theta}_i^{(1,2)} = 1 - exp(-exp(-6{,}018 + 3{,}178X_{2i}^{(1)} + 5{,}954\text{índice-}1_i^{(1)})) \tag{8-1}$$

La figura **8-5** muestra el desplazamiento de la curva ROC de los valores predichos del modelo $\eta^{(1,2)}$ "hacia arriba y a la izquierda", es decir, muy próxima al punto (0,1) y por tanto un área casi máxima, lo que representa el caso óptimo en que las distribuciones de las personas con MC y sin MC se encuentran bien diferenciadas.

El área bajo la curva resultante de la curva ROC anterior fue de $AUC = 97{,}61\,\%$, la cual sostiene la buena capacidad discriminante que el modelo ofrece para clasificar individuos con MC y sin MC. El intervalo de confianza bootstrap del 95 % para el AUC fue $\langle 94{,}83\,\% - 99{,}58\,\%\rangle$, es decir, a partir del método bootstrap con 1000 repeticiones, se tiene un 95 % de confianza que el modelo $\eta^{(1,2)}$ proporcione un muy buen poder discriminante para clasificar a los individuos con y sin MC.

Figura 8-5.: Curva ROC de los valores ajustados del modelo $\eta^{(1,2)}$



Luego, el valor umbral obtenido por medio del índice Youden sugiere que el modelo debe clasificar a una persona con MC si la probabilidad predicha supera el valor 0,24. Por lo tanto la regla de decisión para determinar a una personas con MC viene dada por la variable aleatoria $Z^{(1,2)}$

$$Z^{(1,2)} = \begin{cases} 1 & \text{si } \hat{\theta}_i^{(1,2)} \geq 0{,}24 \\ 0 & \text{si } \hat{\theta}_i^{(1,2)} < 0{,}24 \end{cases} \tag{8-2}$$

Para $i = \{1, 2, ..., 230\}$.

Es decir, en caso de que la probabilidad predicha al incluir la información correspondiente en el modelo sea mayor o igual a 0,24, indicará que la persona tiene MC.

La medida de exactitud o concordancia que ofrece el punto de corte fue $AC = 97{,}39\,\%$, es decir, con la información de la muestra original el 97,39 % de los individuos con MC y sin MC fueron clasificados acertadamente por el modelo. El modelo presentó alta capacidad discrimintante, ya que el valor de exactitud fue un valor muy cercano a 1.

En la tabla **8-17** se observó las veces y sus respectivas fracciones que el modelo clasificó a los individuos con MC y sin MC de manera correcta e incorrecta.

Tabla 8-17.: Tabla de clasificación con el modelo $\eta^{(1,2)}$

	Prueba genética (Gold Standard)	
Modelo	$Y_1 = 1$	$Y_1 = 0$
$Z^{(1,2)} = 1$	15 (88,24 %)	4 (1,88 %)
$Z^{(1,2)} = 0$	2 (11,76 %)	209 (98,12 %)

Donde el 88,24 % de los individuos fueron clasificados con MC por el modelo $\eta^{(1,2)}$ cuando realmente tenían el trastorno y 98,12 % de los individuos fueron clasificados sin MC cuando realmente no presentaban el trastorno. Mas aún, los errores en la clasificación con éste modelo vienen expresados por la fracción de falsos negativos ($\widehat{FFN}$) y la fracción de falsos positivos ($\widehat{FFP}$), el cual no evidencian ser valores altos (11,76 %) y (1,88 %) respectivamente.

Índices de clasificación obtenidos a partir del método bootstrap

A partir de $B = 1000$ muestras bootstrap obtenidas y mediante la aplicación del algoritmo descrito en la sección 7.9, se obtuvieron las estimaciones bootstrap de la medida de exactitud (AC) y las fracciones, que ofreció el modelo $\eta^{(1,2)}$ ajustado para predecir el trastorno MC.

Tabla 8-18.: Estimaciones bootstrap de la medida de exactitud y fracciones con el modelo $\eta^{(1,2)}$

	Valor	SD_{boot}
$\widehat{AC}_{boot}$	97,38 %	0,0102
$\widehat{FVP}_{boot}$	88,28 %	0,0767
$\widehat{FVN}_{boot}$	98,11 %	0,0093
$\widehat{FFN}_{boot}$	11,72 %	0,0767
$\widehat{FFP}_{boot}$	1,89 %	0,0093

En la tabla **8-18**, se pudo observar que a medida de diferentes extracciones de la muestra con repetición, el modelo siguió conservando su buena capacidad discriminante puesto que el $\widehat{AC}_{boot}$ siguió siendo muy cercano a 1. Ademas, los valores $\widehat{FVP}_{boot}$, $\widehat{FVN}_{boot}$, $\widehat{FFN}_{boot}$ y $\widehat{FFP}_{boot}$ mostraron que el modelo presentó buen poder para clasificar a los individuos con trastorno MC y sin MC. Por otro lado, las estimaciones obtenidas en cada muestra bootstrap no se diferenciaron en gran medida, ya que se obtuvo errores estándar bajos.

8.3.2. Estudio de la capacidad predictiva y clasificatoria de los modelos $\eta^{(2,2)}$ y $\eta^{(2,3)}$ ajustados para dar predicción del trastorno PM

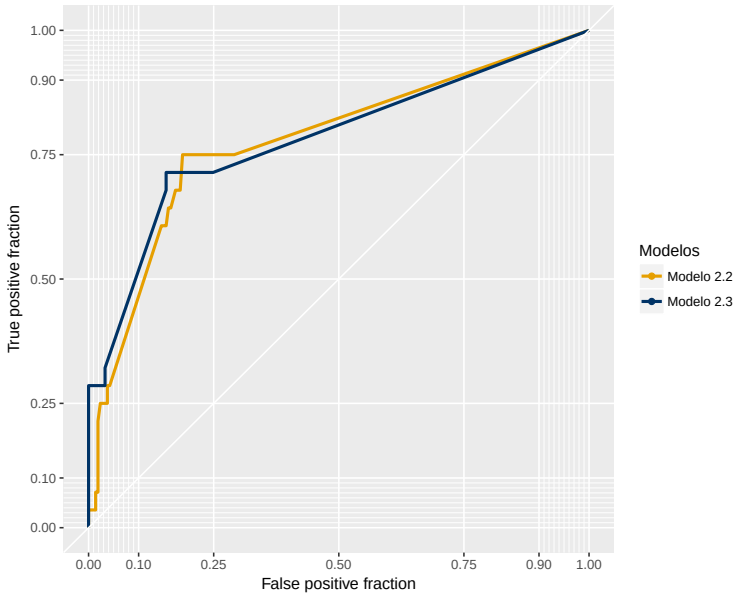
Los valores predichos de los modelos $\eta^{(2,2)}$ y $\eta^{(2,3)}$ son presentados en la siguiente expresión:

$$\hat{\theta}_i^{(2,2)} = 1 - exp(-exp(-2{,}968 + 1{,}857X_{2i}^{(2)} + 2{,}030\text{índice-}1_i^{(2)})) \quad (8\text{-}3)$$

$$\hat{\theta}_i^{(2,3)} = 1 - exp(-exp(-3{,}140 + 1{,}882X_{2i}^{(2)} + 4{,}484\text{índice-}2_i^{(2)})) \quad (8\text{-}4)$$

La curva amarilla y azul presentadas en la figura **8-6**, representan las curvas ROC de los valores predichos de los modelos $\eta^{(2,2)}$ y $\eta^{(2,3)}$ respectivamente. Sus comportamientos evidencian que los modelos no presenten muy alto poder discriminante para clasificar a los individuos con PM y sin PM.

Figura 8-6.: Curvas ROC de los valores predichos de los modelos $\eta^{(2,2)}$ y $\eta^{(2,3)}$



Complementado a lo mencionado anteriormente, la tabla **8-19** muestra que los AUC no fueron muy cercanos a 1, lo cual evidencia que los modelos no presentaron alta capacidad discriminante. Sin embargo, sus intervalos de confianza bootstrap del 95 % mostraron que con un 95 % de confianza los modelos pueden proporcionar máximo 87,32 % y 87,43 % poder discriminante para clasificar los individuos con presencia o ausencia de PM.

Tabla 8-19.: AUC obtenido con los valores predichos de los modelos $\eta^{(2,2)}$ y $\eta^{(2,3)}$

Modelos	AUC $\langle IC(95\,\%)\rangle$
$\eta^{(2,2)}$	$77{,}83\,\%\langle 68{,}10\,\% - 87{,}32\,\%\rangle$
$\eta^{(2,3)}$	$78{,}14\,\%\langle 68{,}06\,\% - 87{,}43\,\%\rangle$

Dada las pocas diferencias observadas en los AUC y sus intervalos de confianza de ambos modelos y teniendo en cuenta que las curvas ROC para los valores predichos de los modelos $\eta^{(2,2)}$ y $\eta^{(2,3)}$ fueron construidas utilizando el mismo conjunto de datos, cabe la duda si solo con el AUC más alto puede escogerse el mejor modelo. Por lo tanto, para comparar cual de los dos modelos proporciona mejor poder discriminante se plantea el siguiente contraste de hipótesis:

$$\begin{cases} H_0: & AUC_1 = AUC_2 \\ H_1: & AUC_2 \neq AUC_2 \end{cases} \tag{8-5}$$

Donde AUC_1 es el área bajo la curva obtenida de la curva ROC para los valores predichos del modelo $\eta^{(2,2)}$ y AUC_2 es el área bajo la curva obtenida de la curva ROC para los valores predichos $\eta^{(2,3)}$.

Luego, como el valor p (0,8774) asociado a la prueba fue mayor al nivel de significancia 5 %, se concluye que no hubo suficiente evidencia muestral para decir que las curvas proporcionan diferente poder discriminante, por lo tanto se obtendrán resultados semejantes si se usa el modelo que incluye el índice de discapacidad obtenido del test WDS2.0 o el índice-2 construido por el método multivariado ACP.
El punto de corte obtenido por medio del índice Youden sugiere que los modelos $\eta^{(2,2)}$ y $\eta^{(2,3)}$ deben clasificar a una persona con PM si la probabilidad predicha supera la siguiente regla de decisión dada por la variable aleatoria $Z^{(2,2)}$ y $Z^{(2,3)}$, para cada modelo respectivamente.

$$Z^{(2,2)} = \begin{cases} 1 & \text{si } \hat{\theta}_i^{(2,2)} \geq 0{,}10 \\ 0 & \text{si } \hat{\theta}_i^{(2,2)} < 0{,}10 \end{cases} \quad Z^{(2,3)} = \begin{cases} 1 & \text{si } \hat{\theta}_i^{(2,3)} \geq 0{,}17 \\ 0 & \text{si } \hat{\theta}_i^{(2,2)} < 0{,}17 \end{cases} \tag{8-6}$$

Para $i = \{1, 2, ..., 241\}$.

Las medidas de exactitud o concordancia que ofreció los puntos de corte para cada modelo presentado en la ecuación 8-6, se pudo observar que aunque las medidas obtenidas no fueron sustancialmente altas, los modelos proporcionaron buena exactitud, puesto que con el 80,50 % y el 82,99 % de los individuos con PM y sin PM fueron clasificados acertadamente por los modelo $\eta^{(2,2)}$ y $\eta^{(2,3)}$, respectivamente.

De acuerdo a los puntos de corte, las tablas **8-20** y **8-21** mostraron el número de veces que los modelos $\eta^{(2,2)}$ y $\eta^{(2,3)}$ respectivamente, clasificaron de manera correcta e incorrecta a los individuos con y sin trastorno PM. El modelo $\eta^{(2,2)}$ ofreció mejor clasificación de los individuos con PM (75,00 %) que el modelo $\eta^{(2,3)}$ (71,43 %). No obstante, para el caso de clasificar personas sin PM el modelo $\eta^{(2,3)}$ proporcionó mejores resultados (84,51 %) que el modelo $\eta^{(2,2)}$ (81,22 %). Sus mínimas diferencias no las hace estadísticamente diferentes, según la prueba de hipótesis anterior.

Tabla 8-20.: Tabla de clasificación con el modelo $\eta^{(2,2)}$

	Prueba genética (Gold Standard)	
Modelo	$Y_2 = 1$	$Y_2 = 0$
$Z^{(2,2)} = 1$	21 (75,00 %)	40 (18,78 %)
$Z^{(2,2)} = 0$	7 (25,00 %)	173 (81,22 %)

Tabla 8-21.: Tabla de clasificación con el modelo $\eta^{(2,3)}$

	Prueba genética (Gold Standard)	
Modelo	$Y_2 = 1$	$Y_2 = 0$
$Z^{(2,3)} = 1$	20 (71,43 %)	33 (15,49 %)
$Z^{(2,3)} = 0$	8 (28,57 %)	180 (84,51 %)

Indices de clasificación obtenidos a partir del método bootstrap

La tabla **8-22** muestra las estimaciones bootstrap de la medida de exactitud AC y las fracciones con los modelos $\eta^{(2,2)}$ $\eta^{(2,3)}$, a partir de $B = 1000$ muestras bootstrap.

Tabla 8-22.: Estimaciones bootstrap de la medida de exactitud y fracciones con los modelos $\eta^{(2,2)}$ y $\eta^{(2,3)}$

	$\eta^{(2,2)}$		$\eta^{(2,3)}$	
Medidas	Valor	SD_{boot}	Valor	SD_{boot}
$\widehat{AC}_{boot}$	80,48 %	0,0256	82,91 %	0,0241
$\widehat{FVP}_{boot}$	74,97 %	0,0830	70,93 %	0,0869
$\widehat{FVN}_{boot}$	81,20 %	0,0267	84,47 %	0,0245
$\widehat{FFN}_{boot}$	25,03 %	0,0830	29,07 %	0,0869
$\widehat{FFP}_{boot}$	18,80 %	0,0267	15,53 %	0,0245

Estos valores muestran que el modelo $\eta^{(2,3)}$ sigue siendo ligeramente mejor en discriminar, ya que el $\widehat{AC}_{boot}$ fue mayor (82,91 %). Sin embargo, en cuestión de clasificar a los individuos con trastorno PM el modelo $\eta^{(2,2)}$ el cual incluye en sus covariables al índice WDS2.0 fue mejor, puesto que la fracción de verdaderos positivos con este modelo fue mayor ($\widehat{FVP}_{boot} = 74{,}97\,\%$) que el obtenido con el modelo $\eta^{(2,3)}$ ($\widehat{FVP}_{boot} = 70{,}93\,\%$). Por otro lado, las estimaciones obtenidas en cada muestra bootstrap no se diferenciaron en gran medida, ya que se obtuvo errores estándar bajos.

9. Conclusiones y recomendaciones

La metodología de los modelos con enlace cloglog, permitieron determinar las variables explicativas que mejor se asociaban con la probabilidad de presentar los trastorno MC y PM, para dar predicción de ellas, a pesar de que el fenómeno separación completa presentado en los modelos 3 y 4 para predecir MC y en el modelo 4 para predecir PM, limitó su ajuste y respectivo análisis. Para este fenómeno, Knoblauch & Maloney (2012) propone ajustar los datos utilizando una función de verosimilitud penalizada que introduce regularización al proceso de ajuste. Por otra parte, para la modelación del trastorno ZG se sugiere el estudio de diferentes transformaciones a los datos, las cuales permitan encontrar un modelo que se ajuste a ellos y dar predicción del trastorno bajo una trasformación. Lo anterior se sugiere para continuación de éste trabajo.

Los resultados obtenidos en la evaluación de la capacidad clasificatoria, mostraron que el modelo seleccionado para predecir el trastorno MC presentó muy buena estructura de clasificación ya que la fracción de verdaderos positivos bootstrap (88,28 %) y verdaderos negativos (98,11 %) fue alta, lo que demostró que el modelo discrimina de forma adecuada las personas con y sin MC, según la información del índice-1 (test WDS2.0) y casos sospechados/confirmados de familia con SXF que se le ingrese al modelo. Además, se tiene en cuenta que la buena propiedad clasificatoria del modelo también se debió a que las personas con este tipo de trastorno presentan problemas de discapacidad muy notorias, a diferencia de las personas con PM, el cual pueden no manifestar problemas de discapacidad y por lo tanto el test WDS2.0 valorarlas sin ningún problema de discapacidad.

Los modelos propuestos para modelar el trastorno PM, mostraron ser menos fuertes para clasificar a los individuos correctamente con PM, evento que se vio evidenciado en la prueba de la capacidad clasificatoria de los modelos. Aunque cabe resaltar que aunque sus mediciones de exactitud no fueron considerablemente altas, éste modelo provee buena capacidad predictiva. Por otro lado, el mejor modelo según el criterio AIC, fue el modelo 3. Sin embargo, al proponer evaluar las capacidades predictivas del modelo 2, ya que éste también presentó buenas características de ajuste, hizo dar cuenta que sus resultados no eran significativamente diferentes y que por lo tanto, es posible hacer uso de uno o de otro para dar predicción del trastorno PM. No obstante, en el desarrollo de la metodología bootstrap, se observó que el segundo modelo el cual incluye en sus covariables al índice obtenido del WDS2.0 daba mejores resultados para clasificar a los individuos con PM.

Los modelos seleccionados en este estudio podrán ser replicables en otra muestra de individuos, por medio de obtención de la información de la pregunta 19 y 20 del cuestionario y bajo los mismos escenarios propuestos en el estudio previo de Lopez & Diaz (2017), construir el índice-2. En el caso del modelo 2 para predecir el trastorno MC, debe accederse al cuestionario WDS2.0 y a partir de obtención de los puntajes obtener el índice-1.

A. Tabla resumen de las medidas de exactitud y fracciones

Tabla A-1.: Tabla resumen de las medidas de exactitud y fracciones

Modelo	Evento éxito	Punto de Corte	AUC	AC	FVP	FVN	FFN	FFP
$\eta^{(1,2)}$	MC	0,24	97,61 %	97,39 %	88,24 %	98,12 %	11,76 %	1,88 %
$\eta^{(2,2)}$	PM	0,10	77,83 %	80,50 %	75,00 %	81,22 %	25,00 %	18,78 %
$\eta^{(2,3)}$	PM	0,17	78,14 %	82,99 %	71,43 %	84,51 %	28,57 %	15,49 %

B. Valores ajustados

B.1. Valores ajustados del modelo $\eta^{(1,2)}$

Individuo	Ajustado	Estado	Pregunta19	PuntajeWDS2
1	0.06	1	1	0.00
2	0.68	1	1	0.50
3	0.05	1	0	0.50
4	0.59	1	1	0.46
5	0.68	1	1	0.50
6	0.77	1	1	0.54
7	0.72	1	1	0.52
8	0.77	1	1	0.54
9	0.92	1	1	0.63
10	0.96	1	1	0.67
11	0.28	1	1	0.29
12	0.68	1	1	0.50
13	0.72	1	1	0.52
14	0.72	1	1	0.52
15	0.72	1	1	0.52
16	0.72	1	1	0.52
17	0.72	1	1	0.52
18	0.00	0	0	0.00
19	0.00	0	0	0.00
20	0.00	0	0	0.00
21	0.00	0	0	0.00
22	0.00	0	0	0.00
23	0.00	0	0	0.00
24	0.00	0	0	0.00
25	0.00	0	0	0.00
26	0.00	0	0	0.00
27	0.00	0	0	0.00
28	0.00	0	0	0.00
29	0.00	0	0	0.00
30	0.06	0	1	0.00
31	0.00	0	0	0.00
32	0.07	0	0	0.56
33	0.00	0	0	0.00
34	0.01	0	0	0.19
35	0.00	0	0	0.00
36	0.00	0	0	0.00
37	0.96	0	1	0.67
38	0.00	0	0	0.00
39	0.00	0	0	0.00
40	0.06	0	1	0.00
41	0.06	0	1	0.00
42	0.00	0	0	0.00
43	0.00	0	0	0.00
44	0.06	0	1	0.00
45	0.55	0	1	0.44
46	0.00	0	0	0.00
47	0.31	0	1	0.31
48	0.12	0	1	0.13
49	0.06	0	1	0.00
50	0.00	0	0	0.00
51	0.12	0	1	0.13
52	0.06	0	1	0.00
53	0.00	0	0	0.00
54	0.06	0	1	0.00
55	0.00	0	0	0.00
56	0.00	0	0	0.00
57	0.00	0	0	0.00
58	0.00	0	0	0.00
59	0.06	0	1	0.00
60	0.00	0	0	0.00
61	0.00	0	0	0.00
62	0.00	0	0	0.00
63	0.01	0	0	0.21
64	0.01	0	0	0.23
65	0.01	0	0	0.15
66	0.00	0	0	0.00
67	0.00	0	0	0.00
68	0.00	0	0	0.00
69	0.06	0	0	0.54
70	0.00	0	0	0.00
71	0.00	0	0	0.02
72	0.00	0	0	0.10
73	0.00	0	0	0.00
74	0.00	0	0	0.00
75	0.00	0	0	0.00
76	0.84	0	1	0.58
77	0.00	0	0	0.00
78	0.00	0	0	0.00
79	0.06	0	1	0.00
80	0.06	0	1	0.00
81	0.06	0	1	0.00
82	0.00	0	0	0.00
83	0.00	0	0	0.00
84	0.01	0	0	0.19
85	0.00	0	0	0.00
86	0.06	0	1	0.00
87	0.00	0	0	0.00
88	0.00	0	0	0.00
89	0.00	0	0	0.00
90	0.00	0	0	0.00
91	0.06	0	1	0.00
92	0.06	0	1	0.00
93	0.06	0	1	0.00
94	0.08	0	1	0.06
95	0.00	0	0	0.00
96	0.00	0	0	0.00
97	0.02	0	0	0.31
98	0.00	0	0	0.00
99	0.00	0	0	0.00
100	0.00	0	0	0.00
101	0.00	0	0	0.00
102	0.00	0	0	0.00
103	0.00	0	0	0.02
104	0.00	0	0	0.06
105	0.00	0	0	0.00
106	0.00	0	0	0.00
107	0.00	0	0	0.10
108	0.00	0	0	0.00
109	0.04	0	0	0.46
110	0.00	0	0	0.00
111	0.00	0	0	0.00
112	0.00	0	0	0.00
113	0.00	0	0	0.00
114	0.00	0	0	0.00
115	0.00	0	0	0.00
116	0.01	0	0	0.23
117	0.04	0	0	0.46
118	0.04	0	0	0.48
119	0.00	0	0	0.00
120	0.00	0	0	0.00
121	0.06	0	1	0.00
122	0.06	0	1	0.00
123	0.00	0	0	0.00
124	0.00	0	0	0.00
125	0.00	0	0	0.02
126	0.00	0	0	0.00
127	0.01	0	0	0.29
128	0.03	0	0	0.42
129	0.00	0	0	0.00
130	0.00	0	0	0.00
131	0.01	0	0	0.25
132	0.00	0	0	0.00
133	0.00	0	0	0.00
134	0.00	0	0	0.00
135	0.00	0	0	0.00
136	0.00	0	0	0.00
137	0.00	0	0	0.00

Individuo	Ajustado	Estado	Pregunta19	PuntajeWDS2
138	0.00	0	0	0.00
139	0.00	0	0	0.00
140	0.00	0	0	0.00
141	0.00	0	0	0.00
142	0.00	0	0	0.00
143	0.00	0	0	0.02
144	0.00	0	0	0.00
145	0.00	0	0	0.00
146	0.00	0	0	0.00
147	0.00	0	0	0.06
148	0.00	0	0	0.00
149	0.00	0	0	0.00
150	0.00	0	0	0.00
151	0.00	0	0	0.00
152	0.00	0	0	0.00
153	0.00	0	0	0.00
154	0.06	0	1	0.00
155	0.00	0	0	0.10
156	0.00	0	0	0.00
157	0.00	0	0	0.00
158	0.00	0	0	0.00
159	0.00	0	0	0.00
160	0.21	0	1	0.23
161	0.00	0	0	0.00
162	0.00	0	0	0.00
163	0.00	0	0	0.00
164	0.12	0	0	0.67
165	0.03	0	0	0.42
166	0.00	0	0	0.00
167	0.00	0	0	0.00
168	0.00	0	0	0.00
169	0.00	0	0	0.00
170	0.00	0	0	0.00
171	0.02	0	0	0.38
172	0.01	0	0	0.23
173	0.00	0	0	0.00
174	0.00	0	0	0.00
175	0.00	0	0	0.00
176	0.00	0	0	0.00
177	0.00	0	0	0.00
178	0.00	0	0	0.00
179	0.00	0	0	0.00
180	0.18	0	1	0.21
181	0.00	0	0	0.00
182	0.06	0	1	0.00
183	0.00	0	0	0.00

Individuo	Ajustado	Estado	Pregunta19	PuntajeWDS2
184	0.00	0	0	0.10
185	0.00	0	0	0.00
186	0.01	0	0	0.21
187	0.00	0	0	0.00
188	0.00	0	0	0.00
189	0.00	0	0	0.00
190	0.06	0	1	0.00
191	0.00	0	0	0.00
192	0.00	0	0	0.00
193	0.00	0	0	0.00
194	0.00	0	0	0.00
195	0.00	0	0	0.00
196	0.00	0	0	0.00
197	0.00	0	0	0.00
198	0.00	0	0	0.00
199	0.00	0	0	0.00
200	0.00	0	0	0.00
201	0.00	0	0	0.00
202	0.00	0	0	0.00
203	0.00	0	0	0.00
204	0.00	0	0	0.00
205	0.00	0	0	0.00
206	0.00	0	0	0.00
207	0.00	0	0	0.00
208	0.00	0	0	0.00
209	0.00	0	0	0.00
210	0.00	0	0	0.00
211	0.00	0	0	0.00
212	0.00	0	0	0.00
213	0.00	0	0	0.00
214	0.00	0	0	0.00
215	0.00	0	0	0.00
216	0.00	0	0	0.00
217	0.00	0	0	0.00
218	0.00	0	0	0.00
219	0.00	0	0	0.00
220	0.00	0	0	0.00
221	0.00	0	0	0.00
222	0.00	0	0	0.00
223	0.00	0	0	0.00
224	0.01	0	0	0.15
225	0.00	0	0	0.00
226	0.06	0	1	0.00
227	0.00	0	0	0.00
228	0.06	0	1	0.00
229	0.00	0	0	0.00
230	0.00	0	0	0.00

B.2. Valores ajustados del modelo $\eta^{(2,2)}$

Individuo	Ajustado	Estado	Pregunta19	PuntajeWDS2
1	0.05	1	0	0.00
2	0.43	1	1	0.27
3	0.78	1	1	0.75
4	0.05	1	0	0.00
5	0.28	1	1	0.00
6	0.05	1	0	0.00
7	0.12	1	0	0.46
8	0.28	1	1	0.00
9	0.28	1	1	0.00
10	0.51	1	1	0.38
11	0.05	1	0	0.00
12	0.05	1	0	0.00
13	0.28	1	1	0.00
14	0.28	1	1	0.00
15	0.28	1	1	0.00
16	0.05	1	0	0.00
17	0.05	1	0	0.00
18	0.14	1	0	0.54
19	0.11	1	0	0.38
20	0.11	1	0	0.38
21	0.45	1	1	0.29
22	0.32	1	1	0.08
23	0.43	1	1	0.27
24	0.42	1	1	0.25
25	0.28	1	1	0.00
26	0.41	1	1	0.23
27	0.28	1	1	0.00
28	0.28	1	1	0.00
29	0.05	0	0	0.00
30	0.05	0	0	0.00
31	0.05	0	0	0.00
32	0.05	0	0	0.00
33	0.05	0	0	0.00
34	0.05	0	0	0.00
35	0.05	0	0	0.00
36	0.05	0	0	0.00
37	0.05	0	0	0.00
38	0.05	0	0	0.00
39	0.05	0	0	0.00
40	0.05	0	0	0.00
41	0.28	0	1	0.00
42	0.05	0	0	0.00
43	0.15	0	0	0.56
44	0.05	0	0	0.00
45	0.07	0	0	0.19
46	0.05	0	0	0.00
47	0.05	0	0	0.00
48	0.72	0	1	0.67
49	0.05	0	0	0.00
50	0.05	0	0	0.00
51	0.28	0	1	0.00
52	0.28	0	1	0.00
53	0.05	0	0	0.00
54	0.05	0	0	0.00
55	0.28	0	1	0.00
56	0.55	0	1	0.44
57	0.05	0	0	0.00
58	0.46	0	1	0.31
59	0.35	0	1	0.13
60	0.28	0	1	0.00
61	0.05	0	0	0.00
62	0.35	0	1	0.13
63	0.28	0	1	0.00
64	0.05	0	0	0.00
65	0.28	0	1	0.00
66	0.05	0	0	0.00
67	0.05	0	0	0.00
68	0.05	0	0	0.00
69	0.05	0	0	0.00
70	0.28	0	1	0.00
71	0.05	0	0	0.00
72	0.05	0	0	0.00
73	0.05	0	0	0.00
74	0.08	0	0	0.21
75	0.08	0	0	0.23
76	0.07	0	0	0.15
77	0.05	0	0	0.00
78	0.05	0	0	0.00
79	0.05	0	0	0.00
80	0.14	0	0	0.54
81	0.05	0	0	0.00
82	0.05	0	0	0.02
83	0.06	0	0	0.10
84	0.05	0	0	0.00
85	0.05	0	0	0.00
86	0.05	0	0	0.00
87	0.66	0	1	0.58
88	0.05	0	0	0.00
89	0.05	0	0	0.00
90	0.28	0	1	0.00
91	0.28	0	1	0.00
92	0.28	0	1	0.00
93	0.05	0	0	0.00
94	0.05	0	0	0.00
95	0.07	0	0	0.19
96	0.05	0	0	0.00
97	0.28	0	1	0.00
98	0.05	0	0	0.00
99	0.05	0	0	0.00
100	0.05	0	0	0.00
101	0.05	0	0	0.00
102	0.28	0	1	0.00
103	0.28	0	1	0.00
104	0.28	0	1	0.00
105	0.31	0	1	0.06
106	0.05	0	0	0.00
107	0.05	0	0	0.00
108	0.09	0	0	0.31
109	0.05	0	0	0.00
110	0.05	0	0	0.00
111	0.05	0	0	0.00
112	0.05	0	0	0.00
113	0.05	0	0	0.00
114	0.05	0	0	0.02
115	0.06	0	0	0.06
116	0.05	0	0	0.00
117	0.05	0	0	0.00
118	0.06	0	0	0.10
119	0.05	0	0	0.00
120	0.12	0	0	0.46
121	0.05	0	0	0.00
122	0.05	0	0	0.00
123	0.05	0	0	0.00
124	0.05	0	0	0.00
125	0.05	0	0	0.00
126	0.05	0	0	0.00
127	0.08	0	0	0.23
128	0.12	0	0	0.46
129	0.13	0	0	0.48
130	0.05	0	0	0.00
131	0.05	0	0	0.00
132	0.28	0	1	0.00
133	0.28	0	1	0.00
134	0.05	0	0	0.00
135	0.05	0	0	0.00
136	0.05	0	0	0.02
137	0.05	0	0	0.00

Individuo	Ajustado	Estado	Pregunta19	PuntajeWDS2
138	0.09	0	0	0.29
139	0.11	0	0	0.42
140	0.05	0	0	0.00
141	0.05	0	0	0.00
142	0.08	0	0	0.25
143	0.05	0	0	0.00
144	0.05	0	0	0.00
145	0.05	0	0	0.00
146	0.05	0	0	0.00
147	0.05	0	0	0.00
148	0.05	0	0	0.00
149	0.05	0	0	0.00
150	0.05	0	0	0.00
151	0.05	0	0	0.00
152	0.05	0	0	0.00
153	0.05	0	0	0.00
154	0.05	0	0	0.02
155	0.05	0	0	0.00
156	0.05	0	0	0.00
157	0.05	0	0	0.00
158	0.06	0	0	0.06
159	0.05	0	0	0.00
160	0.05	0	0	0.00
161	0.05	0	0	0.00
162	0.05	0	0	0.00
163	0.05	0	0	0.00
164	0.05	0	0	0.00
165	0.28	0	1	0.00
166	0.06	0	0	0.10
167	0.05	0	0	0.00
168	0.05	0	0	0.00
169	0.05	0	0	0.00
170	0.05	0	0	0.00
171	0.41	0	1	0.23
172	0.05	0	0	0.00
173	0.05	0	0	0.00
174	0.05	0	0	0.00
175	0.18	0	0	0.67
176	0.11	0	0	0.42
177	0.05	0	0	0.00
178	0.05	0	0	0.00
179	0.05	0	0	0.00
180	0.05	0	0	0.00
181	0.05	0	0	0.00
182	0.11	0	0	0.38
183	0.08	0	0	0.23
184	0.05	0	0	0.00
185	0.05	0	0	0.00
186	0.05	0	0	0.00
187	0.05	0	0	0.00

Individuo	Ajustado	Estado	Pregunta19	PuntajeWDS2
188	0.05	0	0	0.00
189	0.05	0	0	0.00
190	0.05	0	0	0.00
191	0.40	0	1	0.21
192	0.05	0	0	0.00
193	0.28	0	1	0.00
194	0.05	0	0	0.00
195	0.06	0	0	0.10
196	0.05	0	0	0.00
197	0.08	0	0	0.21
198	0.05	0	0	0.00
199	0.05	0	0	0.00
200	0.05	0	0	0.00
201	0.28	0	1	0.00
202	0.05	0	0	0.00
203	0.05	0	0	0.00
204	0.05	0	0	0.00
205	0.05	0	0	0.00
206	0.05	0	0	0.00
207	0.05	0	0	0.00
208	0.05	0	0	0.00
209	0.05	0	0	0.00
210	0.05	0	0	0.00
211	0.05	0	0	0.00
212	0.05	0	0	0.00
213	0.05	0	0	0.00
214	0.05	0	0	0.00
215	0.05	0	0	0.00
216	0.05	0	0	0.00
217	0.05	0	0	0.00
218	0.05	0	0	0.00
219	0.05	0	0	0.00
220	0.05	0	0	0.00
221	0.05	0	0	0.00
222	0.05	0	0	0.00
223	0.05	0	0	0.00
224	0.05	0	0	0.00
225	0.05	0	0	0.00
226	0.05	0	0	0.00
227	0.05	0	0	0.00
228	0.05	0	0	0.00
229	0.05	0	0	0.00
230	0.05	0	0	0.00
231	0.05	0	0	0.00
232	0.05	0	0	0.00
233	0.05	0	0	0.00
234	0.05	0	0	0.00
235	0.07	0	0	0.15
236	0.05	0	0	0.00
237	0.28	0	1	0.00
238	0.05	0	0	0.00
239	0.28	0	1	0.00
240	0.05	0	0	0.00
241	0.05	0	0	0.00

B.3. Valores ajustados del modelo $\eta^{(2,3)}$

Individuo	Ajustado	Estado	Pregunta19	PuntajeWDS2
1	0.04	1	0	0.00
2	0.99	1	1	0.27
3	1.00	1	1	0.75
4	0.04	1	0	0.00
5	0.25	1	1	0.00
6	0.04	1	0	0.00
7	0.81	1	0	0.46
8	0.25	1	1	0.00
9	0.25	1	1	0.00
10	0.25	1	1	0.38
11	0.04	1	0	0.00
12	0.04	1	0	0.00
13	0.25	1	1	0.00
14	0.25	1	1	0.00
15	0.25	1	1	0.00
16	0.04	1	0	0.00
17	0.04	1	0	0.00
18	0.04	1	0	0.54
19	0.23	1	0	0.38
20	0.26	1	0	0.38
21	0.99	1	1	0.29
22	1.00	1	1	0.08
23	0.99	1	1	0.27
24	0.99	1	1	0.25
25	0.25	1	1	0.00
26	0.52	1	1	0.23
27	0.25	1	1	0.00
28	0.25	1	1	0.00
29	0.04	0	0	0.00
30	0.04	0	0	0.00
31	0.04	0	0	0.00
32	0.04	0	0	0.00
33	0.04	0	0	0.00
34	0.04	0	0	0.00
35	0.04	0	0	0.00
36	0.04	0	0	0.00
37	0.04	0	0	0.00
38	0.04	0	0	0.00
39	0.04	0	0	0.00
40	0.04	0	0	0.00
41	0.25	0	1	0.00
42	0.04	0	0	0.00
43	0.10	0	0	0.56
44	0.04	0	0	0.00
45	0.05	0	0	0.19
46	0.04	0	0	0.00
47	0.04	0	0	0.00
48	0.25	0	1	0.67
49	0.04	0	0	0.00
50	0.04	0	0	0.00
51	0.25	0	1	0.00
52	0.25	0	1	0.00
53	0.04	0	0	0.00
54	0.04	0	0	0.00
55	0.25	0	1	0.00
56	0.31	0	1	0.44
57	0.04	0	0	0.00
58	0.31	0	1	0.31
59	0.25	0	1	0.13
60	0.25	0	1	0.00
61	0.04	0	0	0.00
62	0.25	0	1	0.13
63	0.25	0	1	0.00
64	0.04	0	0	0.00
65	0.25	0	1	0.00
66	0.04	0	0	0.00
67	0.04	0	0	0.00
68	0.04	0	0	0.00
69	0.04	0	0	0.00
70	0.25	0	1	0.00
71	0.04	0	0	0.00
72	0.04	0	0	0.00
73	0.04	0	0	0.00
74	0.05	0	0	0.21
75	0.11	0	0	0.23
76	0.04	0	0	0.15
77	0.04	0	0	0.00
78	0.04	0	0	0.00
79	0.04	0	0	0.00
80	0.05	0	0	0.54
81	0.04	0	0	0.00
82	0.04	0	0	0.02
83	0.04	0	0	0.10
84	0.04	0	0	0.00
85	0.04	0	0	0.00
86	0.11	0	0	0.00
87	0.31	0	1	0.58
88	0.04	0	0	0.00
89	0.04	0	0	0.00
90	0.25	0	1	0.00
91	0.25	0	1	0.00
92	0.25	0	1	0.00
93	0.04	0	0	0.00
94	0.04	0	0	0.00
95	0.04	0	0	0.19
96	0.04	0	0	0.00
97	0.25	0	1	0.00
98	0.04	0	0	0.00
99	0.04	0	0	0.00
100	0.04	0	0	0.00
101	0.04	0	0	0.00
102	0.25	0	1	0.00
103	0.25	0	1	0.00
104	0.25	0	1	0.00
105	0.25	0	1	0.06
106	0.04	0	0	0.00
107	0.04	0	0	0.00
108	0.10	0	0	0.31
109	0.04	0	0	0.00
110	0.04	0	0	0.00
111	0.04	0	0	0.00
112	0.04	0	0	0.00
113	0.04	0	0	0.00
114	0.10	0	0	0.02
115	0.04	0	0	0.06
116	0.04	0	0	0.00
117	0.04	0	0	0.00
118	0.04	0	0	0.10
119	0.04	0	0	0.00
120	0.05	0	0	0.46
121	0.04	0	0	0.00
122	0.04	0	0	0.00
123	0.04	0	0	0.00
124	0.04	0	0	0.00
125	0.04	0	0	0.00
126	0.04	0	0	0.00
127	0.11	0	0	0.23
128	0.10	0	0	0.46
129	0.10	0	0	0.48
130	0.04	0	0	0.00
131	0.04	0	0	0.00
132	0.25	0	1	0.00
133	0.25	0	1	0.00
134	0.04	0	0	0.00
135	0.04	0	0	0.00
136	0.11	0	0	0.02
137	0.04	0	0	0.00

Individuo	Ajustado	Estado	Pregunta19	PuntajeWDS2
138	0.04	0	0	0.29
139	0.04	0	0	0.42
140	0.04	0	0	0.00
141	0.04	0	0	0.00
142	0.04	0	0	0.25
143	0.04	0	0	0.00
144	0.04	0	0	0.00
145	0.04	0	0	0.00
146	0.04	0	0	0.00
147	0.04	0	0	0.00
148	0.04	0	0	0.00
149	0.04	0	0	0.00
150	0.04	0	0	0.00
151	0.04	0	0	0.00
152	0.04	0	0	0.00
153	0.04	0	0	0.00
154	0.04	0	0	0.02
155	0.04	0	0	0.00
156	0.04	0	0	0.00
157	0.04	0	0	0.00
158	0.10	0	0	0.06
159	0.04	0	0	0.00
160	0.04	0	0	0.00
161	0.04	0	0	0.00
162	0.04	0	0	0.00
163	0.04	0	0	0.00
164	0.04	0	0	0.00
165	0.25	0	1	0.00
166	0.04	0	0	0.10
167	0.04	0	0	0.00
168	0.04	0	0	0.00
169	0.04	0	0	0.00
170	0.04	0	0	0.00
171	0.49	0	1	0.23
172	0.04	0	0	0.00
173	0.04	0	0	0.00
174	0.04	0	0	0.00
175	0.10	0	0	0.67
176	0.05	0	0	0.42
177	0.04	0	0	0.00
178	0.04	0	0	0.00
179	0.04	0	0	0.00
180	0.04	0	0	0.00
181	0.04	0	0	0.00
182	0.49	0	0	0.38
183	0.49	0	0	0.23
184	0.04	0	0	0.00
185	0.04	0	0	0.00
186	0.04	0	0	0.00
187	0.04	0	0	0.00

Individuo	Ajustado	Estado	Pregunta19	PuntajeWDS2
188	0.04	0	0	0.00
189	0.04	0	0	0.00
190	0.04	0	0	0.00
191	0.31	0	1	0.21
192	0.04	0	0	0.00
193	0.25	0	1	0.00
194	0.04	0	0	0.00
195	0.05	0	0	0.10
196	0.04	0	0	0.00
197	0.05	0	0	0.21
198	0.04	0	0	0.00
199	0.04	0	0	0.00
200	0.04	0	0	0.00
201	0.25	0	1	0.00
202	0.04	0	0	0.00
203	0.04	0	0	0.00
204	0.05	0	0	0.00
205	0.04	0	0	0.00
206	0.04	0	0	0.00
207	0.04	0	0	0.00
208	0.04	0	0	0.00
209	0.04	0	0	0.00
210	0.04	0	0	0.00
211	0.04	0	0	0.00
212	0.04	0	0	0.00
213	0.04	0	0	0.00
214	0.04	0	0	0.00
215	0.04	0	0	0.00
216	0.04	0	0	0.00
217	0.04	0	0	0.00
218	0.04	0	0	0.00
219	0.04	0	0	0.00
220	0.04	0	0	0.00
221	0.04	0	0	0.00
222	0.04	0	0	0.00
223	0.04	0	0	0.00
224	0.04	0	0	0.00
225	0.04	0	0	0.00
226	0.04	0	0	0.00
227	0.04	0	0	0.00
228	0.04	0	0	0.00
229	0.04	0	0	0.00
230	0.04	0	0	0.00
231	0.04	0	0	0.00
232	0.04	0	0	0.00
233	0.04	0	0	0.00
234	0.04	0	0	0.00
235	0.10	0	0	0.15
236	0.04	0	0	0.00
237	0.25	0	1	0.00
238	0.04	0	0	0.00
239	0.25	0	1	0.00
240	0.04	0	0	0.00
241	0.04	0	0	0.00

Bibliografía

Agresti, A. (2015), *Foundations of linear and generalized linear models*, John Wiley & Sons.

Avella, B., Oliva, M. et al. (2010), Comparación del análisis factorial múltiple (AFM) y del análisis en componentes principales para datos cualitativos (Prinqual), en la construcción de índices/Comparison between multiple factor analysis (MFA) and principal component analysis for qualitative data (Prinqual) methods for derivation of indices, PhD thesis, Universidad Nacional de Colombia.

Biggers, A. (2016), 'Everything You Need to Know About Fragile X Syndrome', `https://www.healthline.com/health/fragile-x-syndrome#overview1`. [Web; accedido el 12-10-2017].

Burgueño, M., García Bastos, J. & González Buitrago, J. (1995), 'Las curvas roc en la evaluación de las pruebas diagnósticas', *Medicina clínica* **104**(17), 661–670.

Cabrera, L. (2012), '¿Qué es la discapacidad?', `http://200.33.14.34:1033/archivos/pdfs/Var_104.pdf`. [Web; accedido el 12-10-2017].

Calabrese, R., Osmetti, S. A. et al. (2011), 'Generalized extreme value regression for binary rare events data: an application to credit defaults', *Bulletin of the International Statistical Institute LXII, 58th Session of the International Statistical Institute* pp. 5631–5634.

Cayuela, L. (2009), 'Modelos lineales generalizados (glm)', *Materiales de un curso del R del IREC* .

Cedillo, I. S., Gutiérrez, A. D. & de la Teja Ángeles, E. (2014), 'Aspectos estomatológicos en el síndrome del x frágil. revisión de la literatura y presentación de un caso clínico', *Revista odontológica mexicana* **18**(4), 236–240.

Cerda, J. & Cifuentes, L. (2012), 'Uso de curvas roc en investigación clínica: Aspectos teórico-prácticos', *Revista chilena de infectología* **29**(2), 138–141.

Coffee, B., Keith, K., Albizua, I., Malone, T., Mowrey, J., Sherman, S. L. & Warren, S. T. (2009), 'Incidence of fragile x syndrome by newborn screening for methylated fmr1 dna', *The American Journal of Human Genetics* **85**(4), 503–514.

Efron, B. (1992), Bootstrap methods: another look at the jackknife, *in* 'Breakthroughs in statistics', Springer, pp. 569–593.

Escobar, L. (2006), 'Indicadores sintéticos de calidad ambiental: un modelo general para grandes zonas urbanas', *Eure (Santiago)* **32**(96), 73–98.

Everett, B. (2013), *An introduction to latent variable models*, Springer Science & Business Media.

Faraway, J. J. (2014), *Linear models with R*, CRC press.

Flórez, O. M. & Rincón, W. A. (2012), 'Modelo logit y probit un caso de aplicación', *Universidad Santo Tomas* .

Franco, A. S. (2000), 'Descubierta la causa del retardo mental', `https://aupec.univalle.edu.co/informes/diciembre00/retardo.html`. [Web; accedido el 02-10-2017].

Galbraith, J., Moustaki, I., Bartholomew, D. J. & Steele, F. (2008), *Analysis Multivarite Social Science Data*, CRC Press.

Gil Abreu, S. N. (2014), 'Bootstrap en poblaciones finitas'.

Gunderson, E. P., Jacobs, D. R., Chiang, V., Lewis, C. E., Tsai, A., Quesenberry, C. P. & Sidney, S. (2009), 'Childbearing is associated with higher incidence of the metabolic syndrome among women of reproductive age controlling for measurements before pregnancy: the cardia study', *American journal of obstetrics and gynecology* **201**(2), 177–e1.

Gutiérrez, R. B. (2003), *Validación de supuestos en el modelo de regresión*, Universidad del Valle, Escuela de Ingeniería Industrial y Estadística, Facultad de Ingeniería.

Hajian-Tilaki, K. (2013), 'Receiver operating characteristic (roc) curve analysis for medical diagnostic test evaluation', *Caspian journal of internal medicine* **4**(2), 627.

Hanley, J. A. & McNeil, B. J. (1983), 'A method of comparing the areas under receiver operating characteristic curves derived from the same cases.', *Radiology* **148**(3), 839–843.

Hardin, J. W., Hilbe, J. M. & Hilbe, J. (2007), *Generalized linear models and extensions*, Stata press.

Kidd, S. A., Lachiewicz, A., Barbouth, D., Blitz, R. K., Delahunty, C., McBrien, D., Visootsak, J. & Berry-Kravis, E. (2014), 'Fragile x syndrome: a review of associated medical problems', *Pediatrics* **134**(5), 995–1005.

Knoblauch, K. & Maloney, L. T. (2012), *Modeling psychophysical data in R*, Vol. 32, Springer Science & Business Media.

Kotz, S. & Nadarajah, S. (2000), *Extreme value distributions: theory and applications*, World Scientific.

López de Ullibarri, G. & Pita, S. (1998), 'Curvas roc', *Cad Aten Primaria* **5**(4), 229–35.

Lopez, J. E. & Diaz, J. D. (2017), 'Desarrollo de un índice que permita clasificar individuos con problemas de discapacidad a partir de la pregunta 20 formulada en el cuestionario "genética médica poblacional"', *Universidad del Valle* .

Lorente, L. G. (2017), 'Síndrome x frágil: fisiopatología y posibilidades actuales de diagnóstico y tratamiento'.

Martín, A. C. (2015), 'Síndrome de X Frágil, diagnóstico e intervención', `http://www.bekiapadres.com/articulos/sindrome-x-fragil-diagnostico-intervencion/`. [Web; accedido el 03-10-2017].

Martorell, L. (2010), 'Síndrome X-Frágil', `https://www.sjdhospitalbarcelona.org/sites/default/files/u1/Para_profesionales/Laboratorio/09_laboratori_diagnostic_sindrome_x_fragil.pdf`. [Web; accedido el 03-10-2017].

Mauchant, D., Rice, K. D., Riley, M. A., Leber, D., Samarov, D. & Forster, A. L. (2011), *Analysis of three different regression models to estimate the ballistic performance of new and environmentally conditioned body armor*, US Department of Commerce, National Institute of Standards and Technology.

Mayoral, M. A. M. & Socuéllamos, J. M. (2001), *Modelos lineales generalizados*, Universidad Miguel Hernández.

McCullagh, P. & Nelder, J. A. (1989), *Generalized Linear Models, no. 37 in Monograph on Statistics and Applied Probability*, Chapman & Hall,.

Mingarro Castillo, M., Ejarque Doménech, I., García Moreno, A., Portilla, A. & Miguel, L. (2017), 'Síndrome del cromosoma x frágil', *Revista Clínica de Medicina de Familia* **10**(1), 54–57.

Minguez, M. (2006), 'Síndrome de x-fragil libro de consulta para familias y profesionales', *Madrid: Ed Real Patronato sobre Discapacidad. Madrid* pp. 113–116.

Montgomery, D. C. D. C., Peck, E. A. & Vining, G. G. (2001), *Introducción al análisis de regresión lineal*, number 04; QA278. 2,. M6.

Organización Mundial de la Salud (2001), 'Clasificación internacional del funcionamiento de la discapacidad y de la salud: Cif: versión abreviada'.

Organización Mundial de la Salud (2010), *Measuring health and disability: Manual for WHO disability assessment schedule WHODAS 2.0*, World Health Organization.

Payan, C., Saldarriaga, W., Isaza, C. & Alzate, A. (2001), 'Estudio de foco endémico de retardo mental en ricaurte, valle', *Acta Biológica Colombiana* **6**(2), 88.

Pérez Fernández, S. (2015), 'Estimación de la curva roc acumulativa/dinámica'.

Pita Fernández, S. & Pértegas Díaz, S. (2003), 'Pruebas diagnósticas: Sensibilidad y especificidad', *Cad Aten Primaria* **10**, 120–4.

R Core Team (2017), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

Saldarriaga, W., Tassone, F., Yuriko, L., Forero, J. V., Ayala, S. & Hagerman, R. (2014), 'Sindrome de X Frágil', `http://colombiamedica.univalle.edu.co/index.php/comedica/article/view/1810/2577`. [Web; accedido el 12-10-2017].

Salech, F., Mery, V., Larrondo, F. & Rada, G. (2008), 'Estudios que evalúan un test diagnóstico: interpretando sus resultados', *Revista médica de Chile* **136**(9), 1208–1208.

Solanas, A. & Olivera, V. (1992), 'Bootstrap: fundamentos e introducción a sus aplicaciones', *Anuario de psicología/The UB Journal of psychology* (55), 143–154.

Torréns, J. I., Sutton-Tyrrell, K., Zhao, X., Matthews, K., Brockwell, S., Sowers, M. & Santoro, N. (2009), 'Relative androgen excess during the menopausal transition predicts incident metabolic syndrome in mid-life women: Swan', *Menopause (New York, NY)* **16**(2), 257.

Valle Benavides, A. R. d. (2017), 'Curvas roc (receiver-operating-characteristic) y sus aplicaciones'.

Van der Paal, B. (2014), 'A comparison of different methods for modelling rare events data', *Ghent University* .

Zarzosa, P. (1996), 'Aproximación a la medición del bienestar social', *Valladolid: Secretario de Publicaciones* .

Zepeda, M. & Vásquez, A. (2012), 'Aplicación de la clasificación internacional del funcionamiento, de la discapacidad y de la salud en estudios de prevalencia de discapacidad en las américas', *Organización Panamericana de la Salud* .