



Propuesta metodológica para la construcción de una escala de medición de Riesgo para contraer ITS / VIH-SIDA.

Lorena Ibarra Mera

Universidad del Valle

Facultad de Ingeniería, Escuela de Estadística

Santiago de Cali, Colombia

2018

Propuesta metodológica para la construcción de una escala de medición de Riesgo para contraer ITS / VIH-SIDA.

Lorena Ibarra Mera

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

Director:

PhD. Jose Rafael Tovar Cuevas

Codirector(a):

Ps.MSc. Linda Teresa Orcasita

Universidad del Valle

Facultad de Ingeniería, Escuela de Estadística

Santiago de Cali, Colombia

2017

Dedicatoria

El presente trabajo va dirigido a todas las personas que en el transcurso de mi carrera, me hicieron más fuerte con sus enseñanzas, actitudes y retos. A todos aquellos que creyeron en mí como mi familia, el profesor Jaime Mosquera por su regaño motivacional, mi mamá que siempre fue una fuerte guía para culminar este trabajo de grado y mi padre que me tuvo paciencia.

Agradecimientos

Le agradezco a DIOS por darme la oportunidad de cursar mi pregrado, conocer nuevas personas y aprender de la estadística. A mis padres por proporcionarme las herramientas económicas para asistir a mis clases y cumplir con mis responsabilidades.

A los docentes que conocí gracias por compartirme su conocimiento y enseñarme de la vida en el tiempo que compartí dentro de la universidad. A mi único amigo dentro de la carrera Jhon Lopez, gracias por el tiempo compartido, las vivencias y todos los retos que en su momento superamos no solo académicos sino emocionales.

Resumen

En este trabajo se encuentra desarrollada una propuesta metodológica estadística la cual plantea un indicador de riesgo para estudiantes universitarios por medio del Análisis factorial. Se utilizó información obtenida por medio de un cuestionario llamado Reconociendo mi salud sexual reproductiva II”, el cual se le aplicó a estudiantes de tres universidades de la ciudad de Cali, en total se trabajó con 760 respuestas de estudiantes y 30 ítems, algunos originales del cuestionario y otros compuestos de varias preguntas del instrumento. Luego de analizar la información descriptivamente y por medio del análisis factorial se obtuvo 11 componentes las cuales fueron agrupadas y formaron un solo indicador y una única puntuación para cada estudiante al ser evaluado. A estos valores se les realiza una transformación la cual permite que sean interpretados los resultados con mayor facilidad ante personas que desconozcan la parte numérica estadística. Con esta base, se plantean otras dos propuestas de indicador las cuales varían en la matriz utilizada para evaluar a los estudiantes, dichas propuestas no funcionan tan bien como la original así que se decide tomar el indicador inicial como finalista para definir si un estudiante está o no en riesgo de contraer ITS y VIH/SIDA. Finalmente se realiza una simulación donde se pone a prueba el indicador planteado pero bajo un escenario un poco más controlado en el cual el indicador mejora y es capaz de puntuar correctamente a estudiantes que si se encuentran en riesgo y deben ser intervenidos.

Palabras clave: VIH/SIDA, variable latente, ITS, índice, Análisis factorial, variable cualitativa, validez, curva ROC, Distribución, Riesgo, Condón)

Abstract

In this work, a statistical methodological proposal is developed which raises a risk indicator for university students through Factorial Analysis. We used information obtained through a questionnaire called Recognizing my reproductive sexual health II”, which was applied to students of three universities in the city of Cali, in total it worked with 760 student responses and 30 items, some original of the questionnaire and other compounds of several questions of the instrument. After analyzing the information descriptively and by means of factorial analysis, 11 components were obtained which were grouped and formed a single indicator and a single score for each student to be evaluated. At these values ??a transformation is made which allows them to be interpreted the results with greater ease before people who do not know the statistical numerical part. With this base, two other indicator proposals are proposed, which vary in the matrix used to evaluate the students. These proposals do

not work as well as the original, so it is decided to take the initial indicator as a finalist to define if a student is or not at risk of contracting STIs and HIV / AIDS. Finally, a simulation is carried out where the proposed indicator is put to the test but under a more controlled scenario in which the indicator improves and is able to correctly score students who are at risk and should be intervened.

Keywords: VIH/SIDA, latent variable, ITS, index, Factor analysis, qualitative variable, validity, ROC curve, Distribution, Risk, Condom

Contenido

Resumen	vii
Lista de Figuras	xii
Lista de Tablas	1
1 Introducción	3
1.1 Planteamiento del Problema	3
1.2 Justificación	5
1.3 Objetivos	7
1.4 Antecedentes	8
2 Marco teórico	12
2.1 Marco teórico conceptual	12
2.1.1 Entidades Reguladoras	14
2.2 Marco teórico estadístico	16
2.2.1 Variable Latente	16
2.3 Teoría de los TEST	17
2.3.1 Validez	17
3 Metodología	22
3.1 Ítems y análisis descriptivo	23
3.2 Aplicación Análisis Factorial (AF)	29
3.2.1 Análisis factorial AF	30
3.2.2 Aplicación Análisis factorial	31
3.2.3 Pesos factoriales	31
3.2.4 Valores Propios	32
3.2.5 Componentes	32
3.3 Indicador	34
3.3.1 Puntuación por componente	35
3.4 Otros rangos para el indicador	38
3.4.1 Datos originales	38
3.4.2 Normalización de datos observados	38
3.5 Escalas de riesgo	38

3.6	Puntos de corte	41
4	Resultados	43
4.1	Análisis descriptivo	43
4.2	Indicador	44
4.2.1	Curva ROC-Punto de corte	45
5	Uso del Indicador de Riesgo	50
5.1	Estudio de Simulación	50
5.2	Aplicación del índice de riesgo por universidad	52
5.2.1	Universidad 1	52
5.2.2	Universidad 2	53
5.2.3	Universidad 3	53
5.2.4	Universidad 4	54
6	Conclusiones y recomendaciones	56
6.1	Conclusiones	56
6.2	Recomendaciones	57
	Bibliografía	58

Lista de Figuras

1-1	Representación del VIH en el mundo	5
1-2	ITS a nivel mundial según OMS	5
2-1	Curva ROC	20
3-1	Comparación de las agrupaciones utilizando datos imputados vs datos reales	24
3-2	Indice IR- Matriz x	39
3-3	Indice IR- Matriz X	40
3-4	Indice IR- Matriz X'	40
4-1	Curva ROC Indicador 1	46
4-2	Curva ROC Indicador 2 y 3	48
5-1	Curva ROC Indice 1 simulado	51

Lista de Tablas

3-1	Pesos factoriales	32
3-2	Resumen niveles de riesgo, según matriz de respuestas utilizadas	41
4-1	Uso del preservativo vs Tener sexo siempre al salir	44
4-2	Retribución económica o material a cambio de sexo y, Uso del condón.	44
4-3	Clasificación Indicador 1	47
4-4	Clasificación Indicador 2 y 3	48
5-1	Clasificación Indicador 1 datos simulados	51

Declaración

Me permito afirmar que he realizado el presente Trabajo de Grado de manera autónoma y con la única ayuda de los medios permitidos y no diferentes a los mencionados en el propio trabajo. Todos los pasajes que se han tomado de manera textual o figurativa de textos publicados y no publicados, los he reconocido. Ninguna parte del presente trabajo se ha empleado en ningún otro tipo de Tesis o Trabajo de Grado.

Igualmente declaro que los datos utilizados en este trabajo están protegidos por las correspondientes cláusulas de confidencialidad.

Santiago de Cali, 13.Julio.2018

Lorena Ibarra Mera

1 Introducción

1.1. Planteamiento del Problema

En el mundo cada día, más de un millón de personas contraen una infección de transmisión sexual(ITS)OMS (2016), se estima que, anualmente, unos 75 millones de personas contraen alguna de las siguientes cuatro infecciones de transmisión sexual(IST): clamidiasis, gonorrea, sífilis o tricomoniasis. Según cálculos de la OMS y el ONUSIDA, a finales de 2016 había en el mundo unos 36,7 millones de personas infectadas por VIH. Ese mismo año, contrajeron la infección 1,8 millones de personas, y un millón murió por causas relacionadas con el VIH OMS (2017b) . Las cifras son tan alarmantes que se considera como un problema de salud pública mundial llevando a la creación de ONUSIDA dirigida por Dr. Michele Sidibé en la cual, fue decretado para el año 2001 la *"Declaración de compromiso en la lucha contra el VIH/SIDA"* de Naciones Unidas) (2000). Today-CONDOMS (n.d.) expone los factores que aumentan la posibilidad de adquirir una ITS:

"Estos factores incluyen los comportamientos sexuales, la percepción de riesgo (a veces las personas pueden pensar que es algo que no les va a tocar o que no les puede pasar), y las habilidades personales para tomar decisiones acertadas. También hay costos y obstáculos que en ocasiones impiden tener comportamientos que ayudan a reducir el riesgo individual de adquirir enfermedades de transmisión sexual." Today-CONDOMS (n.d.).

Profesores de tres universidades de la ciudad de Cali realizaron una investigación acerca de las conductas, comportamientos, conocimientos y creencias de los jóvenes universitarios sobre la sexualidad y la prevención de ITS y VIH. Este estudio hace parte de un proyecto llamado *"TRAMAS"*, en el cual se quiso observar las redes que componen dichos comportamientos y actitudes en jóvenes universitarios.

Para esto se tuvo en cuenta estudiantes de las tres universidades de la ciudad de Cali a las que pertenecían los docentes. En el estudio se aplicó un cuestionario llamado *"Reconociendo mi salud sexual y reproductiva II"*, el cual, contiene 119 ítems, ninguno de carácter abierto y la manera de aplicación fue vía web, anónima y no obligatoria. Luego de que se discutieran muchos aspectos de la investigación se encontró que desde un inicio no hubo un acompañamiento adecuado por parte de un estadístico, que los psicólogos participantes de la investigación hicieron ajustes a cuestionarios anteriores según su criterio y además no tuvieron en cuenta aspectos como la validez de dicho instrumento.

Debido a esto se encontró que el test sólo cumple con validez de contenido realizada

por expertos y no cumple con la validez de constructo, además que la caracterización en la escala de medición no fue definida y era necesario plantear una propuesta metodológica de escala, que permita cuantificar el riesgo de contraer VIH e ITS en jóvenes universitarios. A consecuencia de lo anterior se sugiere que en pruebas donde es primordial la exactitud del estudio se tenga en cuenta la validez de la prueba, la cual incorpora propiedades relevantes dentro del estudio como el buen funcionamiento del test es decir que, tenga la capacidad de reflejar la realidad de lo que se está observando. La necesidad de proponer una escala de medición con métodos estadísticos parte de que tanto las personas que trabajan con salud en temas de ITS y VIH como las entidades universitarias y la sociedad en general, tienen la necesidad de un indicador que permita observar las conductas de universitarios que, los exponen a contraer alguna enfermedad con mayor o menor riesgo.

Lo más importante en el desarrollo del indicador es que a pesar de la metodología empleada para su creación, se espera que los resultados que este arroje concuerden con la realidad de los estudiantes y lo esperado psicológicamente según las conductas y hábitos de los universitarios.

1.2. Justificación

Según la OMS (2017b)(Organización mundial de la Salud) en su centro de prensa en noviembre de 2013 realizan una nota descriptiva (110) sobre las ITS, en esta nota se indica que las ITS son causadas por más de 30 bacterias, virus y parásitos todos diferentes. La manera predominante de propagación es por contacto sexual ya sea vaginal, anal u oral. Otro dato importante es que las ITS se encuentran dentro de las cinco categorías principales por la cual un adulto consulta, además de que cada día, más de un millón de personas contraen una infección de transmisión sexual.

La OMS en su centro de prensa en una nota descriptiva (N° 360) publicada en noviembre de 2015, explica detalladamente la manera en que el VIH ataca el sistema inmunitario y debilita todo el sistema de defensas en una persona, especialmente el de infecciones y algunos tipos de cáncer. Cuando se habla de SIDA es porque el VIH se encuentra en su fase más avanzada, donde aparecen tipos de cáncer, infecciones o manifestaciones clínicas de alta complejidad.



Figura 1-1: Representación del VIH en el mundo



Figura 1-2: ITS a nivel mundial según OMS

Considerando que son cifras alarmantes, problemáticas a nivel mundial de salud pública y que no todas las personas toman conciencia de sus prácticas sexuales se considera relevante

el estudio de comportamientos y prácticas sexuales en estudiantes universitarios de cuatro universidades de la ciudad de Cali y la creación de una escala de medición o índice de Riesgo de contraer VIH/SIDA e ITS. Se plantea para dicho desarrollo técnicas psicométricas, teoría del Test y en especial métodos multivariados, ya que estos nos permiten trabajar con variables cualitativas y determinar las variables relevantes y el número de variables adecuadas para tener el mejor acercamiento de nivel de riesgo en que se encuentra cada persona según sus conductas o características. La idea de medir el riesgo de contraer una ITS o VIH/SIDA en estudiantes universitarios es a causa de que estos en cifras mundiales y estudios han demostrado que son la población con mayor casos de infección y se consideran vulnerables. Dicha vulnerabilidad se debe a factores de riesgo relacionados con el sexo a temprana edad, drogadicción, desigualdad social y género. A nivel mundial se estima que 11,8 millones de jóvenes entre 14 y 24 años viven en la actualidad con VIH/SIDA y, cada día, al rededor de 6000 jóvenes de esas edades se infectan con el VIH y solo una parte de ellos sabe que esta infectado.

Los métodos multivariados se espera permitan analizar las diferentes formas en que una persona presente un mayor riesgo de contagio, es decir, por medio de combinaciones de características que presente una persona se pueda determinar qué tan vulnerable esté a contraer, desarrollar o sea portador de alguna ITS o VIH. Ante la situación planteada y la preocupación de que cada día las cifras de contagio son mayores se plantea la siguiente pregunta de investigación, la cual se espera ayude a responder los objetivos de esta investigación ¿Cuál es la mejor combinación de ítems del instrumento "Reconociendo mi salud sexual y reproductiva II" que permite acercarse al nivel más real de riesgo para contraer ITS o VIH?

1.3. Objetivos

Objetivo General

Proponer una metodología estadística para la construcción de una escala de medición que permita medir el riesgo de contraer ITS y VIH/SIDA en estudiantes universitarios.

Objetivos Específicos

- Desarrollar una metodología estadística para la construcción de una escala que permita generar una expresión de riesgo.
- Sintetizar la escala de riesgo a través de un índice que cuantifique la expresión de riesgo.
- Clasificar a los estudiantes de acuerdo con su nivel de riesgo.

1.4. Antecedentes

En seguida serán citados algunos trabajos, artículos e informes académicos nacionales e internacionales sobre el VIH/SIDA e ITS los cuales están relacionados con los objetivos del trabajo a desarrollar y permiten tener ideas sobre como se debe abordar el problema planteado.

Cada antecedente será presentado de forma resumida donde se expone el objetivo de este y los resultados obtenidos de acuerdo a la metodología utilizada.

Chacon & Corrales (2009) : La educación sexual dirigida a las personas durante la adolescencia no es solo un deber, sino también un derecho que garantiza la prevención de embarazos no deseados e infecciones de transmisión sexual, además el disfrute de su sexualidad plena que forma parte de la salud física y mental de las personas.

USA2016William (2015): *Homonegative Attitudes and Risk Behaviors for HIV and Other Sexually Transmitted Infections Among Sexually Active Men in the United States*

Se examinaron las asociaciones entre las actitudes homonegativas entre los comportamientos de riesgo de hombres estadounidenses sexualmente activos. Se utilizó la Encuesta Nacional de Crecimiento Familiar 2006-2010 multivariable y modelos de regresión logística para examinar las actitudes homonegativas en relación con el uso del condón.

Se encontró que entre los hombres que tuvieron relaciones sexuales con hombres, las actitudes homonegativas se asociaron con menores probabilidades de uso del condón durante el sexo anal con mujeres (antes del año anterior) y las pruebas de ITS en el último año. Entre los hombres que tuvieron relaciones sexuales con hombres y mujeres, las actitudes homonegativas se asociaron con menores probabilidades de uso de condones durante el sexo vaginal y el sexo con hombres, con 4 o más parejas y con la prueba del VIH. Entre los hombres que tuvieron relaciones sexuales con mujeres, las actitudes homonegativas se asociaron con menores probabilidades de uso de condones durante el sexo vaginal y el sexo con hombres (antes del año anterior), pruebas de VIH alguna vez y contraer herpes, virus del papiloma humano o sífilis. Al final llegaron a la conclusión de que las actitudes homonegativas pueden promover la adquisición y transmisión del VIH / ITS entre hombres sexualmente activos de todas las orientaciones sexuales.

Evelynne (2015): *ClustVarLV: An R Package for the Clustering of Variables Around Latent Variables*

En este trabajo , el método CLV se representa como la agrupación de variables latentes alrededor de Variables enfoque (CLV). Teniendo en cuenta sucesivamente diferentes tipos de situaciones , los criterios de CLV subyacentes son detallados y las diversas funciones del

paquete se ilustran mediante estudios de casos reales .

Giron (2013):*HIV sexual risk behaviors in youth 15-24 years of age in Cali, Colombia: Do differences exist among neighborhoods?*

Se buscó establecer diferencias en comportamientos sexuales de riesgo para VIH entre jóvenes de 15 a 24 años de dos áreas de la Ciudad de Cali. Se realizó un estudio transversal en jóvenes de esas edades y un muestreo probabilístico bietápico.

En el estudio transversal se encontró que los distritos 13, 15 y 20, la prevalencia de relaciones sexuales sin protección en los últimos 12 meses y la prevalencia de relaciones sexuales con dos o más parejas fue del 70 %; y el 38 % de los jóvenes tuvo relaciones sexuales bajo los efectos del alcohol. En ambas áreas, la intención se relacionó positivamente con los comportamientos de riesgo. Al analizar factores sociodemográficos se observó que las intenciones y creencias aumentan la oportunidad de mostrar estos comportamientos y que, el efecto de estos factores varía según el distrito. Finalmente hacen énfasis en que existe una alta prevalencia de conductas de riesgo para el VIH relacionadas con factores sociodemográficos, intenciones y creencias que justifican intervenciones apropiadas para las realidades locales.

Elkin (2010): Este documento presenta la estimación del Indicador de Calidad de Vida (ICV) para los hogares del departamento de Antioquia. La metodología estadística empleada se basa en el uso de los métodos de Cuantificación de Variables Cualitativas y el Análisis No Lineal de Componentes Principales. El objetivo de este estudio es analizar, basados en la nueva Encuesta de Calidad de Vida realizada en el año 2009, la evolución de las condiciones de vida urbanas y rurales del departamento con respecto a las halladas en el año 2007 y determinar si han cambiando significativamente.

ONU (1989): La comisión de Estadística de las Naciones Unidas aceptó las directrices preliminares sobre indicadores sociales en su decimonoveno período de sesiones, celebrado en 1976. En aquella ocasión la comisión decidió publicar y difundir directrices estadísticas para la realización de dichos estudios.

Barrios (2002): Durante 1999 y 2000 se recogió información del Hospital de Algeciras correspondiente a los ingresos hospitalarios del periodo 1997-1998. Se estudiaron las variables número de ingresos, mortalidad, número de reingreso, número de consultas externas, índice case mix, número de estancias e índice funcional. Se Aplicó la técnica de análisis de componentes principales (ACP), y se utilizó la matriz de correlaciones R.

McCutcheon (2002): Los modelos de clases latentes tuvieron dos percepciones muy importantes por parte de Clifford Clogg entre (1981-1995). Una de estas percepciones fue desde el punto formal de la estadística relacionada con la estimación de los modelos de clase latente y los avances realizados. De esta manera Clifford manifestó que la biometría, la econometría y la estadística tiene lazos muy estrechos en la modelización de variables latentes. En esta edición se demostró la importancia y fácil aplicación del modelo de clase latente, ayudando a diversos problemas de investigación reales.

José Arturo Granados (2009): Debido a la cultura conservadora de México, quisieron realizar un estudio con el fin de describir la relación que existe entre hombres homosexuales y las situaciones homofóbicas a las que estos se enfrentan en el momento de incrementar el riesgo de contraer o transmitir VIH/SIDA. Se realizó un estudio cualitativo, descriptivo, correlacional, no experimental y transversal, dicho estudio fue con información obtenida de 19 jóvenes homosexuales universitarios de la universidad de la Ciudad de México.

Teodoro Barros (2001): En Ecuador se quiso implementar un modelo educativo de prevención primaria de enfermedades de transmisión sexual y VIH/SIDA debido a la alta tasa de incidencia de personas infectadas. Dicho modelo es basado en los conocimientos, actitudes y prácticas sexuales de los adolescentes. El modelo se aplicó a jóvenes entre 12 y 15 años de Santo Domingo de los Colorados, se tuvo un grupo de intervención y un grupo control en los cuales se observó si existían diferencias significativas antes y después de la intervención con la prueba de la x^2 y, se validó un modelo multifactorial de prevención.

Velásquez (2005): Se quiso estudiar el efecto que tenía un CD "Planeta Riesgo Xero" en los conocimientos actitudes y prácticas relacionadas a las ITS y VIH/SIDA de adolescentes de colegios de Lima. El estudio se realizó en una muestra aleatoria de los alumnos del cuarto y quinto año de secundaria de 26 colegios estatales y particulares. Participaron del estudio 454 adolescentes de 14 a 17 años, a quienes se les aplicó una encuesta de conocimientos, Actitudes y Prácticas antes y después de la exposición al CD educativo. La evaluación de la intervención educativa en los colegios se realizó con un diseño pre experimental probabilístico, bietápico. Para medir el grado de diferencia de conocimientos antes de la aplicación del CD y luego de la intervención, se utilizó la prueba de McNemar (utilizado para diseños pareados y con variables categóricas dicotómicas). Se contrastó la siguiente hipótesis nula: Hipótesis nula(H_0): No existe cambio con relación a los conocimientos y

actitudes sobre ITS y VIH/sida de los alumnos seleccionados entre antes y después de haberse implementado el CD "Planeta Riesgo Xero".

y **Espada (2009)**: El objetivo de este estudio fue analizar las conductas sexuales de riesgo en jóvenes universitarios de la provincia de Alicante y su relación con el consumo de sustancias. La muestra estuvo compuesta por 339 estudiantes universitarios, con un rango de edad entre 18 y 29 años. Para determinar si existe una relación significativa entre el consumo de alcohol y el uso de preservativo en el coito vaginal y el consumo de cannabis y uso de preservativo en el coito oral, se uso una prueba χ^2 con test no paramétricos los cuales definían si eran significativas dichas diferencias.

2 Marco teórico

En este capítulo se encuentran los conceptos médicos, psicológicos y estadísticos relacionados con la investigación. Por otra parte se expone la metodología estadística empleada para cumplir con los objetivos del trabajo de grado con un desarrollo ordenado y orientado.

2.1. Marco teórico conceptual

A continuación se exponen los conceptos de mayor relevancia dentro de la investigación. Estos conceptos ayudaran a comprender la importancia de crear un indicador de riesgo y ayudarán a interpretar correctamente los resultados y conclusiones expuestas en los próximos capítulos.

Según la OMS (2017a)(Organización Mundial de la Salud)

VIH/SIDA

El virus de la inmunodeficiencia humana (VIH) ataca el sistema inmunitario y debilita los sistemas de defensa contra las infecciones y contra determinados tipos de cáncer. A medida que el virus destruye las células inmunitarias e impide el normal funcionamiento de la inmunidad, la persona infectada va cayendo gradualmente en una situación de inmunodeficiencia.

La fase más avanzada de la infección por el VIH es el Síndrome de inmunodeficiencia adquirida o sida que, en función de la persona, puede tardar de 2 a 15 años en manifestarse.

ITS(Infecciones de transmisión sexual)

Entre los más de 30 virus, bacterias y parásitos que se sabe se transmiten por contacto sexual, ocho se han vinculado a la máxima incidencia de enfermedades de transmisión sexual. De esas 8 infecciones, 4 son actualmente curables, a saber, la sífilis, la gonorrea, la clamidiasis y la tricomoniasis. Las otras 4 hepatitis B, virus del herpes simple (HSV o herpes), VIH y virus del papiloma humano (VPH)? son infecciones virales incurables,

aunque existen tratamientos capaces de atenuar o modificar los síntomas o la enfermedad. Las ITS se propagan predominantemente por contacto sexual, incluidos el sexo vaginal, anal y oral. Una persona puede tener una ITS sin manifestar síntomas de enfermedad. Los síntomas comunes de las ITS incluyen flujo vaginal, secreción uretral o ardor en los hombres, úlceras genitales y dolor abdominal

Factor de riesgo

Un factor de riesgo es cualquier rasgo, característica o exposición de un individuo que aumente su probabilidad de sufrir una enfermedad o lesión de salud OMS (2016). A continuación se exponen los factores inmediatos que influyen en el riesgo individual de adquirir una infección de transmisión sexual(ITS).

1. Relaciones sexuales a una edad temprana
2. Mayor número de parejas
3. Parejas de alto riesgo(Personas con altas probabilidades de haber tenido muchas parejas)
4. Aumento de la frecuencia de las relaciones sexuales y ciertas prácticas sexuales.
5. La falta de la circuncisión de la pareja masculina.
6. El uso de duchas vaginales.
7. La falta de uso de condón.

Vías de transmisión

Se presentan las formas más comunes de contraer ITS y VIH.

1. **Sexual:**

- * Por contacto sexual no protegido (pene-ano, pene-vagina pene-boca) con una persona con VIH.

2. **Sanguínea:**

- * Por transfusiones de sangre o sus derivados (plasma, plaquetas) que tienen virus.
- * Por trasplante de órganos con VIH.
- * Por compartir agujas/jeringas en personas usuarias de drogas inyectables (UDI).

3. Perinatal:

- * Una mujer embarazada con VIH puede transmitir el virus al bebé en cualquier momento del embarazo.
- * Durante el parto, a través del canal vaginal por el contacto del bebé con las secreciones vaginales potencialmente infectadas.
- * Por medio de la leche materna (lactancia).

Prevención

Proceso para proporcionar a las poblaciones los medios necesarios para mejorar la salud y ejercer un mayor control sobre la misma, mediante la intervención de los determinantes de la salud y la reducción de la inequidad. Esto se desarrolla fundamentalmente a través de los siguientes campos: formulación de política pública, creación de ambientes favorables a la salud, fortalecimiento de la acción y participación comunitaria, desarrollo de actitudes personales saludables y la reorientación de los servicios de salud Minsalud (2016).

Después de tener una idea clara de que es prevención pasaremos a exponer los métodos de prevención para no contraer ni propagar las ITS ó VIH Hernandez & Lopez (2008).

- Fidelidad
- Abstinencia
- Sexo seguro
- Sexo protegido

Condón

Today-CONDOMS (n.d.) CONDOMS define el condón como una barrera de protección que cubre el pene durante las relaciones sexuales. Está hecho de múltiples materiales entre los que se destaca el látex por su resistencia y elasticidad. Tiene el fin de impedir el paso y flujo del semen antes, durante o después de la eyaculación. Los condones hechos de tejido animal como piel de cordero, no son recomendables para protegernos contra infecciones de transmisión sexual (ITS) ya que, según Giménez (2016) los virus y bacterias logran pasar por los "agujeros" naturales de estos preservativos, impidiendo únicamente el paso de espermatozoides.

2.1.1. Entidades Reguladoras

A nivel nacional y mundial existen entidades encargadas de vigilar y promover estrategias para la prevención y control de las ITS y el VIH/SIDA.

Ministerio de Salud Colombiano: En el año 2013 el Ministerio de salud y Protección Social realizó una extensa revisión normativa sobre todos los temas de salud sexual y reproductiva. Las siguientes normas se consideran relevantes para el conocimiento y manejo del tema:

1. **Decreto 1543 de 1997.** Reglamenta el manejo de la infección por el Virus de Inmunodeficiencia Humana (VIH), el Síndrome de la Inmunodeficiencia Adquirida (SIDA) y las otras Enfermedades de Transmisión Sexual
2. **Ley 972 de 15/07/2005.** Adopta normas para mejorar la atención por parte del Estado colombiano a la población que padece de enfermedades ruinosas o catastróficas, especialmente el VIH/Sida.
3. **Circular 63 de 2006.** Cobertura de servicios de salud y la obligatoriedad para la realización de las pruebas diagnósticas y confirmatorias para VIH.
4. **Resolución 2338 de 2013.** Establece directrices para facilitar el acceso al diagnóstico de la infección por VIH y otras Infecciones de Transmisión Sexual - ITS y para el entrenamiento en pruebas rápidas de VIH, sífilis y otras ITS.
5. **Resolución 5592 de 2015.** Actualiza integralmente el Plan de Beneficios en Salud con cargo a la Unidad de Pago por Capitación-UPC del Sistema General de Seguridad Social en Salud ?SGSSS? y se dictan otras disposiciones. Incluye todo lo requerido para la atención integral del VIH/Sida y otras ITS.

OMS(Organización Mundial de la Salud):

La OMS es una organización a nivel mundial la cual tiene dentro de sus principios la salud como un estado de completo bienestar físico, mental y social, y no solamente la ausencia de afecciones o enfermedades. Además para la OMS la salud es uno de los derechos fundamentales de todo ser humano sin distinción de raza, religión, ideología política o condición económica o social.

ONUSIDA): La organización de las naciones unidas contra el VIH/SIDA como una "*Crisis mundial - Acción mundial*" , conformada por los jefes de estado y gobiernos o representantes de estos, reunidos en las Naciones Unidas del 25 al 27 de junio de 2001 con ocasión del vigésimo sexto período extraordinario de sesiones de la Asamblea General, celebrado ,como cuestión de urgencia, de conformidad con la resolución 55/13, de 3 de noviembre de 2000, para examinar y hacer frente al problema del VIH/SIDA en todos

sus aspectos, así como para lograr que todos se comprometan a mejorar la coordinación e intensificar las actividades nacionales, regionales e internacionales para combatirlo de forma integral.

2.2. Marco teórico estadístico

A continuación se presentan las definiciones de los conceptos y teoría estadística que se aplicó a lo largo del desarrollo de este trabajo.

2.2.1. Variable Latente

En la enciclopedia de Springer Alex C. Michalos (2014) se define como variable latente a aquellas cualidades que no se miden directamente, sino que se deducen solo de la covariación observada entre un conjunto de variables observables (Tabachnick y Fidell, 2001). Los análisis multivariados se utilizan para identificar las dimensiones latentes (p. Ej., Análisis de factores, análisis discriminante, Correlación canónica), que se definen por los conjuntos de pesos asignados a las variables observadas. Las dimensiones latentes son los elementos esenciales de interés ya que representan los principales constructos causales de variables hipotéticas y no directamente medibles (Borsboom, Mellenbergh y van Heerden, 2003).

Escala de Intervalo

Cuando a partir de una relación de orden, si además las diferencias entre las clases son asimismo ordenables (la diferencia entre cualquier par de números es mayor, igual o menor que la diferencia entre cualquier otro par de números), entonces se puede definir una magnitud que las ordene respecto a un atributo. Es decir existe una operación o propiedad de la distancia y es posible establecer una unidad de medida. Es decir, se pueden determinar las diferencias entre dos mediciones a partir de una unidad de medida, por tanto, podemos establecer intervalos iguales entre los valores de las variables. Las escalas nominal y la ordinal, no permiten establecer una distancia (amplitud) entre las categorías o clases. En el caso de las variables latentes las cuales son creadas a partir de variables observadas se pueden crear intervalos con la escala final de esta, con la anotación de que se pueda perder parte de la información al intentar transformar nuevamente de escala los resultados.

Análisis factorial

Es una técnica estadística de reducción de datos con la menor pérdida de información que sea posible. (El requisito, generalmente, es que se conserve al menos el 50 % de la información inicialmente contenida en los datos empíricos).

El análisis de factores se desarrolló por primera vez en la psicología para identificar variables latentes no observadas, como la inteligencia general, entre una serie de medidas. Estas variables latentes se derivan como combinaciones de las variables medidas. El método ahora es ampliamente utilizado en las ciencias sociales para identificar patrones generales en las relaciones entre un conjunto de variables, generalmente exhibidas a través de sus coeficientes de correlación. A pesar de que el Análisis Factorial tiene muchos puntos en común con el análisis de componentes principales existe una diferencia que es la de interés y es la búsqueda de explicar la estructura de las covarianzas entre las variables. De la misma manera que en el método de Componentes Principales es necesario que las variables originales no estén incorrelacionadas ya que no habría nada para explicar de éstas.

Análisis en Componentes Principales

El Análisis de Componentes Principales (ACP) es una técnica estadística de síntesis de la información, o reducción de la dimensión (número de variables). Es decir, ante un banco de datos con muchas variables, el objetivo será reducirlas a un menor número perdiendo la menor cantidad de información posible. Los nuevos componentes principales o factores serán una combinación lineal de las variables originales, y además serán independientes entre sí. Gurrea (n.d.)

Elkin (2010) Bajo este contexto, el Análisis de Componentes Principales (ACP) es un procedimiento estadístico para el análisis de datos multivariados que permite:

- Construir indicadores como resúmenes de un conjunto características dadas (Métodos de reducción de dimensión).
- Es útil cuando las variables están relacionadas linealmente y son de tipo cuantitativo.
- Si $X_1, X_2, \dots, X_p$ es el conjunto de características cuantitativas que queremos resumir, entonces el ACP proporciona las p nuevas variables:
$$Y_i = \sum a_{ij} X_j, \text{ componentes.}$$

2.3. Teoría de los TEST

2.3.1. Validez

En el análisis y la construcción del índice es de suma importancia incluir un análisis sobre la coherencia en relación con el objeto de estudio construido. En el momento en que se realizan

mediciones se requiere que éstas cumplan con dos características importantes: validez y fiabilidad. En este orden de ideas una medida es válida si mide aquello que se quiere medir, en el caso de los indicadores empíricos elegidos se espera que proporcionen una representación adecuada, lo mas exacta y auténtica posible, del concepto teórico que miden y que reflejen la realidad del significado teórico. Ya que al hablar de validez se esta refiriendo a los resultados de una prueba o medición y no al procedimiento de medición se exponen los tipos de validez las cuales puedan ser evaluadas en una medición.

- **Interna:** Grado de confianza que se tiene de que los resultados del experimento se interpreten adecuadamente y sean válidos (se logra cuando hay control).
- **Externa:** Posibilidad de generalizar los resultados de un experimento a situaciones no experimentales, así como a otras personas y poblaciones.
- **Contenido:** Se refiere al grado en que un instrumento refleja un dominio específico de contenido de lo que se mide.
- **Criterio:** Se establece al validar un instrumento de medición al compararlo con algún criterio externo que pretende medir lo mismo. Si el criterio se fi ja en el presente de manera paralela, se habla de validez concurrente (los resultados del instrumento se correlacionan con el criterio en el mismo momento o punto de tiempo). Dentro de la validez de criterio se puede encontrar la sensibilidad, la cual es un concepto para instrumentos de clasificación.

4.1. Sensibilidad

La sensibilidad de un test se define como la capacidad que tiene la prueba de clasificar correctamente a los individuos "enfermos" en "enfermos".

$$\text{Sensibilidad} = \frac{\text{verdaderos positivos}}{\text{total de casos positivos}} = \frac{a}{a + c} = \frac{VP}{VP + FN} * 100 \%$$

donde:

a =verdaderos positivos

$a + c$ =total de casos positivos(enfermos)

VP/FN =verdaderos positivos/falsos negativos

Entre más alto sea el valor de la sensibilidad, hay una mejor capacidad en la detección de enfermos por medio de la prueba. Tener una baja sensibilidad en el test produce pérdida de casos que pudieran ser tratados, lo cual es mucho más grave y que, se tendrían pacientes enfermos como sanos los cuales estarían por fuera del tratamiento o contaminando a otros sin saberlo.

- **Constructo:** La más importante, sobre todo desde una perspectiva científica, y se refiere a qué tan exitosamente un instrumento representa y mide un concepto teórico (Grinnell, Williams y Unrau, 2009). A esta validez le concierne en particular el significado del instrumento, esto es, qué está midiendo y cómo opera para medirlo. Integra la evidencia que soporta la interpretación del sentido que poseen las puntuaciones del instrumento (Messick, 1995).

De esta manera a la validez de la medición se le debe establecer su fiabilidad la cual hace referencia a la consistencia de la medición, a su estabilidad, constancia y predictibilidad. Una medida es fiable si su utilización reiterada en el tiempo produce bajo las mismas o similares circunstancias , los mismos resultados.

Especificidad

Es la proporción real de individuos no enfermos clasificados correctamente. Al igual que la sensibilidad entre mayor sea la capacidad de esta para identificar sujetos sanos por medio del test es mejor.

$$Especificidad = \frac{\text{verdaderos negativos}}{\text{total de casos negativos}} = \frac{b}{b + d} = \frac{VN}{VN + FP} * 100 \%$$

donde:

$$\begin{aligned} b &= \text{verdaderos negativos} \\ b + d &= \text{total de casos negativos(sanos)} \\ VN/FP &= \text{verdaderos negativos/falsos positivos} \end{aligned}$$

Curva Roc

Wojtek y David en su publicación sobre la curva ROC para datos continuos como una monografía en estadística y aplicaciones probabilísticas definen la curva ROC como la manera gráfica de mostrar la tasa real de los positivos vs la de los falsos positivos. Es decir que la curva ROC determina el punto de corte en una escala de medición el cual ayuda a diferenciar los sujetos sanos vs enfermos, positivos vs negativos de una manera casi exacta la cual depende de una serie de características o test diagnósticos Hand (2009).

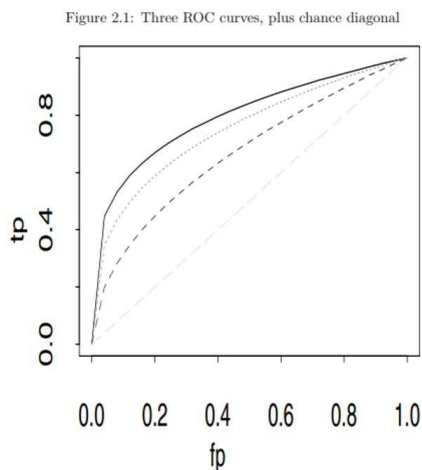


Figura 2-1: Curva ROC

Codificación de respuestas

Luego de recolectar la información(datos) debido al tipo de variables "cualitativas" las categorías de un ítem o pregunta requieren ser codificados con símbolos o números ya que, si esta acción no se realiza no es posible efectuar un análisis detallado, ni la utilización de técnicas estadísticas avanzadas.

Normalización de índices

Se conoce como normalización de índices al proceso de ajustar los valores observados en diferentes escalas a una sola escala común. Para el índice de riesgo propuesto se considero utilizar la escala característica la cual consiste en convertir todos los valores a un rango de $[0,1]$. Este procedimiento es llamado "Normalización basada en la unidad" que a manera general restringe el rango de valores de cualquier conjunto de datos entre dos puntos a y b .

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Software estadístico y librerías

Se utilizó el software estadístico **R-project.org** y la librería Stephane Dray (2017) y *psych*. Estas librerías son herramientas exclusivas para el análisis de datos por medio de métodos multivariantes.

Criterio de Kaiser

Indica que hay que conservar los componentes principales cuyos valores propios son mayores que la unidad, lo cual está estrechamente relacionado con la varianza la cual debe ser vista como la información que recolecta o aporta cada componente. Visto de esta manera aquellas componentes que cumplan este criterio serán las que mayor información aporten o representen con su combinación lineal de variables.

Psicometría

Área fundamental de la psicología que se ocupa de los procedimientos de medición del comportamiento humano Tornimbeni (2008).

De acuerdo con Coombs, Dwes y Tversky (1981) se considera que los papeles fundamentales asignados a la Ciencia es la descripción, explicación y predicción de los fenómenos observables por medio de unas cuantas leyes generales que expresen las relaciones entre las propiedades de los objetos investigados. Las diferencias con los atributos físicos al medir este tipo de variables (psicológicas) se planteó una nueva concepción de medición (Zeller y Carmines 1980) consideraron que se trata de un proceso mediante el cual se enlazan conceptos abstractos (constructos) inobservables directamente, con indicadores empíricos observables directamente (conductas). A este tipo de medición se le suele denominar medición por indicadores, dado que las variables psicológicas no se pueden medir de forma directa, es necesario seleccionar una serie de indicadores que sí pueden ser medidos directamenteOnline (n.d.).

3 Metodología

Este capítulo contiene de manera detallada el proceso que se llevo a cabo en la generación del indicador de riesgo de contraer ITS y VIH/SIDA para estudiantes universitarios. Se debe tener en cuenta que antes de llegar a un solo resultado se dieron muchas vueltas ya que se tuvieron problemas desde la generación y aplicación del cuestionario, hasta el método estadístico más indicado para cumplir con los objetivos.

Se trabajó con una base de datos de tamaño (760x30) la cual se extrajo del archivo original el cual contenía la respuesta de 915 estudiantes a 119 ítems. Las respuestas proporcionadas por los 760 estudiantes eran de escala Likert y dicotómica y, algunos de los ítems fueron conformados a partir de dos o tres preguntas iniciales. La base de datos contenía información relacionada con las prácticas sexuales de los estudiantes, el uso constante del condón, parejas sexuales que había tenido en el último año, si había tenido alguna ITS, entre otras preguntas muy relacionadas con su vida sexual activa además, se creó una pregunta la cual pretendía saber si el estudiante aun no había tenido relaciones sexuales.

La base de datos final con la respuesta de los 760 estudiantes a los 30 ítems se debe a que los 915 estudiantes que registraban inicialmente no contestaron todas las 119 preguntas iniciales lo cual, generó datos faltantes que al momento de ser simulados por métodos de imputación que se explicaran en el próximo capítulo no funcionaron correctamente ya que se distorsionaban los clusters. Las preguntas abiertas no se contemplaron en el estudio, por tal motivo todos los ítems tenían una escala de respuesta. Inicialmente al ordenar la base de datos se encontró que se tenían 915 registros de estudiantes los cuales presentaban "*no respuesta*" en algunos ítems. Debido a esto se intento aplicar la teoría de "Imputación de datos" la cual no funcionó como se esperaba, al comparar las agrupaciones de los datos con información simulada y la real las diferencias eran muy grandes; se tomó la decisión de depurar la base de datos hasta que no quedara ni un dato faltante, lo cual haría que se trabajara solo con información real del comportamiento y hábitos de los estudiantes.

Reducción de dimensionalidad

El cuestionario que se utilizó en los estudiantes universitarios llamado "*Reconociendo mi salud sexual y reproductiva II*" constaba de 119 ítems los cuales estaban divididos en 4 módulos (Información general, Dinámicas sexuales, Salud sexual e infecciones de transmisión sexual, Vida sexual). Un grupo de enfermeras se reunió inicialmente y descartó 68 de los

119 ítems sugiriendo que se trabajara con las 8 variables nuevas las cuales agrupaban la respuesta de dos o tres ítems y las 48 preguntas iniciales que estás bajo criterio medico y psicológico consideraron relevantes.

El desarrollo del indicador de riesgo se empezó a desarrollar con estas 51 variables pero, se encontró que aún era una dimensión muy amplia y los métodos estadísticos como el análisis factorial no lograban realizar agrupaciones reales y concretas sobre las prácticas sexuales de los estudiantes. Debido a esto se creó la matriz de correlaciones entre las 51 variables encontrando que muchas de estas estaban aportando la misma información y, se pidió la revisión de una docente especialista en los comportamientos y practicas sexuales de jóvenes universitarios la cual sugirió que algunos de los 51 ítems no estaban relacionados con el riesgo al que estos estaban expuestos según sus prácticas. Con estos dos criterios se reviso pregunta a pregunta y se tomó la decisión de dejar 30 de los 51 ítems los cuales aportarían la mayor información sobre los estudiantes y su exposición a contraer ITS-VIH/SIDA.

Luego de realizar la última depuración de información se obtuvo una base de datos final con 760 estudiantes los cuales respondieron completamente el cuestionario que contaba con 119 ítems.

Después de que el archivo de datos estuvo listo para ser usado se encontró que el cuestionario tenía una gran cantidad de ítems. Por medio de análisis multivariado (clusters) se intento realizar agrupaciones las cuales contendrían características de comportamientos y hábitos representados en cada ítem. Este intento fue fallido ya que eran demasiados ítems donde la mayoría tenían múltiple opción de respuesta, debido a esto el método no logró agrupar eficientemente en categorías grupos de estudiantes ya que algunos ítems no aportaban información relevante y otros estaban altamente correlacionados. Debido a esto lo que se hizo para minimizar ítems fue tener en cuenta la opinión de algunos expertos que participaron el proceso y podían sugerir los ítems de mayor aporte al analizar el grado de riesgo de los estudiantes universitarios, fue de gran importancia su ayuda ya que, se pasó de tener 119 ítems a trabajar con 30 ítems. Con estos 30 ítems fue mucho más fácil buscar una metodología que ayudara a describir los perfiles de los estudiantes universitarios y permitiera crear un indicador de riesgo.

Finalmente en este capitulo se mostrará un análisis descriptivo de las variables, las transformaciones que se les realizó, la metodología estadística aplicada y la manera en que se conformó el indicador y la escala a cual se llegó.

3.1. Ítems y análisis descriptivo

Teniendo en cuenta lo anterior se trabajó con 30 ítems y 760 respuestas de estudiantes los cuales pertenecían a tres universidades de la región donde dos de estas son de carácter privado y una público. El archivo final de datos contiene ítems modificados, esto quiere decir que

no se encuentran como originalmente se preguntó en el cuestionario. Algunos se convirtieron en preguntas con respuesta dicotómica y otros fueron creados como la unión de dos o tres preguntas y, según la combinación de las posibles respuestas en cada uno se generaba una única respuesta a ese nuevo ítem. De este modo se trabajó con:

- 17 ítems originales los cuales se convirtieron a escala de respuesta de tipo dicotómica.
- 6 nuevos ítems creados a partir de dos o tres preguntas iniciales del cuestionario que presentan escala de respuesta tipo likert.
- 7 ítems originales del cuestionario con escala de respuesta tipo likert.

En el software estadístico R se encuentran dos librerías llamadas *MICE* y *MissingDataGUI* las cuales se complementan y dependiendo del tipo de dato faltante aplican uno de sus modelos o métodos con el cual se generará el valor. *MissingDataGUI* utiliza métodos multivariados dependiendo del tipo de variables para imputar la información; en este caso las variables fueron de tipo cualitativas por lo tanto, el método a emplear fue "mice" el cual va enlazado con la librería *MICE* que contiene el algoritmo MICE (Imputación Múltiple de Especificación Condicional) descrito por Van Buuren y Groothuis-Oudshoorn. Como los métodos empleados para utilizar dicho algoritmo son multivariados se permite aprovechar esta información y generar de acuerdo al método, clusters con los cuales fue posible comparar que tan potente era la imputación de dichos datos y la similitud que se tenía respecto a lo real. La comparación se hizo observando las formaciones que generaba la base de datos con algunas respuestas simuladas según el método de imputación versus las base de datos con el 100 % de las respuestas al cuestionario por parte de cada estudiante. Como se observa las agrupaciones que realiza el método entre los datos imputados y los reales son bastante diferentes, algunas cambian de eje, otras se separan y forman grupos mayores o menores y lo más importante, las características entre estos no coinciden, es decir que no tendremos estudiantes con hábitos y comportamientos similares a lo que realmente estos expresaron si utilizamos el método de imputación.

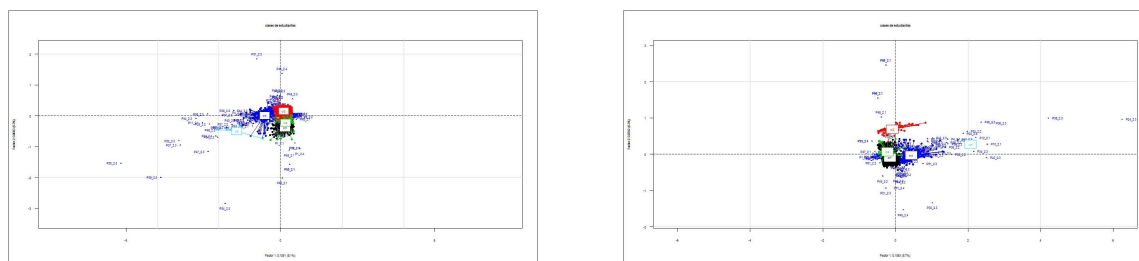


Figura 3-1: Comparación de las agrupaciones utilizando datos imputados vs datos reales

Luego de haber evidenciado la no eficiencia de imputar información (respuestas de algunos estudiantes), se expone propiamente las características de las variables que se utilizaron en

la creación de indicador. De las variables no conformadas, es decir las variables originales utilizadas y seleccionadas pero transformadas a respuesta dicotómica se tiene que éstas representaban en los estudiantes aspectos como:

1. Actividades que realiza en su tiempo libre.
2. Utilización de los servicios de salud en su universidad.
3. Lugares que frecuenta.
4. Hábitos y encuentros sexuales.
5. Salud sexual.

Por otra parte se tienen las 6 variables conformadas a partir de dos o tres ítems las cuales representan:

1. Tipo de riesgo según el tipo de protección y pareja(estable u ocasional).
2. Tipo de riesgo relacionado con las prácticas sexuales.
3. Actividad sexual actual.
4. Virginidad

A continuación se presentan de manera detallada las 30 preguntas que fundamentan las variables con las cuales se trabajó y se creó el indicador de riesgo.

P1 Te gusta más para disfrutar de tu tiempo libre salir solo a tomar cerveza y conocer personas?

Pregunta con respuesta en escala dicotómica.

P2 Te gusta más para disfrutar de tu tiempo libre ir a rumbear solo y conocer personas?

Pregunta con respuesta en escala dicotómica.

P3 En tu universidad venden preservativos?

Pregunta con respuesta en escala dicotómica.

P4 En algún momento has utilizado alguno de los servicios en el área de salud sexual y reproductiva que brinda tu universidad?

Pregunta con respuesta en escala dicotómica.

P5 Para la creación de esta variable se tienen en cuenta la pregunta 63 y 65:

63. ¿Has tenido relaciones sexuales en el último año?

65. ¿En el último mes, más o menos, cuantas veces tuviste relaciones sexuales?

Pregunta con respuesta en escala LIKERT creada donde se podían generar las siguientes respuestas

0. Inactivo: 63. No ? 65. (No tuve y No recuerdo)
1. Menos activo sexualmente: 63. Si ? 65. (No tuve y No recuerdo)
2. Activo sexualmente: 63. Si ? 65. (Entre 1 y 3 veces; Entre 4 y 10 veces; Más de 10 veces)

P6 Alguna vez te has hecho la prueba del VIH?

Pregunta con respuesta en escala dicotómica.

P7 Has hecho pactos de intercambio de sangre con alguien?

Pregunta con respuesta en escala dicotómica.

P8 Cuando una relación se transforma de informal a seria (noviazgo), no es necesario continuar usando preservativo?

Pregunta con respuesta en escala dicotómica.

P9 Con qué frecuencia has tenido tipo de encuentros sexuales como prácticas sexuales en contra de tu voluntad?

Pregunta con respuesta en escala LIKERT donde podían responder:Nunca,Algunas veces,Casi siempre,Siempre,No respondió.

P10 Cuándo sales a rumbar, para pasarla bien, necesitas o te gusta consumir bebidas alcohólicas?

Pregunta con respuesta en escala dicotómica.

P11 Vida sexual activa 4: Ha tenido prácticas sexuales en el último año?

Pregunta creada con respuesta en escala LIKERT donde se podían encontrar combinaciones de respuestas como:

1. Nunca ha tenido prácticas sexuales penetrativas
2. No ha tenido prácticas sexuales penetrativas en el último año
3. Ha tenido prácticas sexuales pero no penetrativas en el último año
4. Ha tenido prácticas sexuales penetrativas pero no en el último año
5. Ha tenido prácticas sexuales penetrativas en el último año

P12 Con qué frecuencia has tenido tipo de encuentros sexuales como prácticas sexuales con intercambio de parejas?

Pregunta con respuesta en escala LIKERT donde podían responder:Nunca,Algunas veces,Casi siempre,Siempre,No respondió.

P13 Has tenido infecciones de transmisión sexual?

Pregunta con respuesta en escala dicotómica.

- P14 Cuándo sales a rumbear, para pasarla bien, necesitas o te gusta consumir sustancias alucinógenas?
Pregunta con respuesta en escala dicotómica.
- P15 Con qué frecuencia has tenido tipo de encuentros sexuales como prácticas sexuales con alguna persona a la que le hayas dado retribución monetaria o material, por ejemplo: matrícula, ropa, celulares, accesorios, zapatos, calificaciones. ?
Pregunta con respuesta en escala LIKERT donde podían responder:Nunca,Algunas veces,Casi siempre,Siempre,No respondio.
- P16 Cuándo sales a rumbear, para pasarla bien, necesitas o te gusta fumar cigarrillos, consumir sustancias alucinógenas,consumir bebidas alcohólicas y tener sexo?
Pregunta con respuesta en escala dicotómica.
- P17 Con qué frecuencia has tenido tipo de encuentros sexuales como prácticas sexuales con más de una persona a la vez?
Pregunta con respuesta en escala LIKERT donde podían responder:Nunca,Algunas veces,Casi siempre,Siempre,No respondio.
- P18 Utilizar un método anticonceptivo oral o inyectado, ha hecho que en adelante no utilices el preservativo o método de barrera con tu pareja?
Pregunta con respuesta en escala dicotómica.
- P19 Con qué frecuencia has tenido tipo de encuentros sexuales como prácticas sexuales con personas que ejercen alguna forma de trabajo sexual?
Pregunta con respuesta en escala LIKERT donde podían responder:Nunca,Algunas veces,Casi siempre,Siempre,No respondio.
- P20 Cuándo sales a rumbear, para pasarla bien, necesitas o te gusta tener sexo?
Pregunta con respuesta en escala dicotómica.
- P21 En tus prácticas sexuales, ¿incluyes o incluirías intercambio de parejas?
Pregunta con respuesta en escala dicotómica.
- P22 En la última relación sexual usaste el preservativo o método de barrera?
Pregunta con respuesta en escala dicotómica.
- P23 Con qué frecuencia has tenido tipo de encuentros sexuales como prácticas sexuales usando alguna sustancias psicoactivas (SPA), para intensificar tu experiencia?
Pregunta con respuesta en escala LIKERT donde podían responder:Nunca,Algunas veces,Casi siempre,Siempre,No respondio.
- P24 En tus prácticas sexuales, ¿incluyes o incluirías la participación en encuentros grupales?
Pregunta con respuesta en escala dicotómica.

P25 Riesgo 3: Prácticas sexuales vs Uso del condón en parejas estables y ocasionales.

No existe riesgo si el estudiante se protege con el condón, independientemente si tiene o no práctica sexual. Por otro lado se da una calificación de 1 si el estudiante no presenta la práctica sexual y a su vez tiene un riesgo mínimo y alto frente al uso del condón. Se destaca con una calificación de 2 si el estudiante presenta un riesgo mínimo en el uso del condón y a su vez presenta alguna práctica sexual. Y por último si el estudiante presenta un riesgo alto frente al uso del condón y presenta algún tipo de práctica sexual se califica como un riesgo 3.

Por lo tanto al sumar las 7 opciones de prácticas sexuales se tiene un rango de calificación de 0 a 21. Se establece el siguiente criterio para la creación de la nueva variable.

- 0 - 0 : No hay riesgo
- De 1 a 7 - 1 : Riesgo mínimo
- De 8 a 14 - 2 : Riesgo alto
- De 15 a 21 - 3 : Riesgo total

P26 Con qué frecuencia has tenido tipo de encuentros sexuales como prácticas sexuales usando alguna bebida alcohólica para intensificar tu experiencia?

Pregunta con respuesta en escala LIKERT donde podían responder: Nunca, Algunas veces, Casi siempre, Siempre, No respondio.

P27 Usas el preservativo u otros métodos de barrera en tus prácticas sexuales?

Pregunta con respuesta en escala dicotómica.

P28 Riesgo 2: Cruce de 66a y 70a (No protección y pareja no estable)

Según las combinaciones de las respuestas se tiene que:

- 1. P70a. Si ? P66a. Si - No hay riesgo
- 2. P70a. No ? P66a. Si - Riesgo mínimo
- 3. P70a. Si ? P66a. No - No hay riesgo
- 4. P70a. No ? P66a. No - Riesgo Total

P29 Riesgo 4: Prácticas sexuales vs Uso de preservativo en parejas estables.

Según el cruce se tienen 4 grupos por orden de riesgo para cada una de las 7 variables generadas:

- 1. 67. Si ? P70a. Si - No hay riesgo

- 2. 67. No ? P70a. Si - No hay riesgo
- 3. 67. Si ? P70a. No - Riesgo total
- 4. 67. No ? P70a. No - Riesgo mínimo

P30 Riesgo 1: Cruce de 64 y 70. Representa el riesgo frente al uso del condón o método de barrera en parejas ocasionales y estables.

- 0. No hay riesgo
- 1. Riesgo mínimo
- 2. Riesgo alto
- 3. Riesgo total

Usando las 30 variables categóricas se realizó un análisis descriptivo el cual se compone de recuentos acerca de la opinión de los estudiantes, cuantos de los que participaron tienen vida sexual activa, cuantos usan protección y la frecuencia con que la usan. Lo otro que se observó con el análisis descriptivo fue la cantidad de individuos que presentan comportamientos y realizan actos que implican riesgo de infección por transmisión sexual y el porcentaje de estudiantes que sostienen relaciones con dos o más parejas sexuales.

Dentro de la metodología estadística se tuvo en cuenta del análisis multivariado: Análisis de correspondencia múltiple (ACM) y Análisis Factorial (AF). Entre estos dos métodos existe una gran diferencia la cual se ve reflejada en la manera que el método trata y asume las variables cualitativas; el análisis de correspondencia múltiple NO tiene en cuenta que exista un orden o una prioridad entre las respuestas es decir que $1 = 2 = 3 = 4 = 5$ mientras que el análisis factorial logra identificar que $1 < 2 < 3 < 4 < 5$ representando un mayor riesgo a los estudiantes en sus respuestas. Por esta razón se utilizó únicamente el análisis factorial AF.

3.2. Aplicación Análisis Factorial (AF)

El principio del análisis factorial es encontrar un subespacio de dimensiones reducidas el cual represente de manera óptima las distancias entre los individuos y la estructura de las correlaciones entre las variables, incluyendo el estudio de los grupos de variables. Ahora con mayor importancia se debe tener en cuenta que intentar representar el riesgo de contraer ITS y VIH/SIDA en estudiantes es buscar la manera de medir una variable latente, la cual como se explicó al principio no es medible cuantitativamente con algún instrumento pero si se puede tener una aproximación de está por medio del tratamiento adecuado de variables observadas que pueden ser tanto cualitativas como cuantitativas o ambas.

3.2.1. Análisis factorial AF

Sean $X_1, \dots, X_{30}$ variables aleatorias generadas a partir de las respuestas dadas por los estudiantes de las tres universidades de la ciudad de Cali a las preguntas del cuestionario. Se quiere encontrar $p+q$ nuevas variables las cuales son llamadas factores comunes(q) y factores únicos(p) , $F_1, \dots, F_q, U_1, \dots, U_p$, tales que:

$$X_1 = \lambda_{11}F_1 + \lambda_{12}F_2 + \dots \lambda_{1q}F_q + U_1$$

...

$$X_p = \lambda_{p1}F_1 + \lambda_{p2}F_2 + \dots \lambda_{pq}F_q + U_p$$

donde se hallarán:

$$F_1, \dots, F_q = q \text{ Factores comunes}$$

$$U_1, \dots, U_p = p \text{ Factores unicos}$$

En el modelo factorial lineal se supone:

- $p < q$ Se debe explicar la relación entre variables con un número reducido de factores en el cual, los " q " factores comunes deben ser menor a los " p " factores únicos hallados. Esto se debe a que se crean p factores únicos conocidos como las componentes las cuales están formadas por la combinación lineal de q factores comunes representados por las preguntas o ítems que conforman la componente.
- Los $q + p$ factores son variables no correlacionadas.

Desglosando un poco más el análisis factorial se presentan a continuación las matrices generadas teniendo en cuenta las 30 variables, la manera en que trabaja el método y como este ayudó a conformar posibles modelos y representaciones que serían importantes en la creación del indicador de riesgo.

En este caso puntual de que no es posible medir el riesgo de contraer ITS y VIH/SIDA de manera directa y, se tienen variables de interés como sus preferencias, hábitos y comportamientos sexuales, se utilizan recursos como codificar las respuestas de los individuos, métodos que generen agrupaciones y tendencias según la similitud entre ellos y modelos lineales con el peso de cada respuesta a una característica del individuo. La variable de interés "*RIESGO DE CONTRAER ITS Y VIH/SIDA*" sería la variable latente de este estudio y, la metodología para encontrar su representación a partir de unas variables observadas es conocido como Análisis Factorial el cual, es un modelo de regresión múltiple que relaciona variables latentes con observadas.

3.2.2. Aplicación Análisis factorial

Se tuvo en cuenta el conjunto de $q = 30$ variables observadas donde $X' = (x_1, x_2, \dots, x_{30})$ las cuales se asume están relacionadas con un número de variables latentes $F_1, F_2, \dots, F_q$, donde $p < 30$, mediante una relación de tipo

$$X_1 = \lambda_{11}F_1 + \lambda_{12}F_2 + \dots \lambda_{1q}F_q + U_1$$

...

$$X_p = \lambda_{p1}F_1 + \lambda_{p2}F_2 + \dots \lambda_{pq}F_q + U_p$$

La cual conforma una matriz

$$x = \Lambda f + u$$

donde

$$\Lambda = \begin{pmatrix} \lambda_{11} & \cdots & \lambda_{1q} \\ \vdots & \ddots & \vdots \\ \lambda_{301} & \cdots & \lambda_{30q} \end{pmatrix}, f = \begin{pmatrix} f_1 \\ \vdots \\ f_q \end{pmatrix}, u = \begin{pmatrix} u_1 \\ \vdots \\ u_q \end{pmatrix}$$

Utilizando métodos computacionales se observa en la matriz de correlación (30x30) que la mayoría de variables están incorrelacionadas lo cual hace posible la aplicación del método y la efectividad de este. La importancia de que las variables se encuentren no correlacionadas se debe a la varianza que cada una de éstas aporta al método convertida en información relevante para la componente y, por esta razón se escoge como primer componente a aquella que contenga la mayor varianza de todas ya que será la más relevante en la interpretación de los resultados.

3.2.3. Pesos factoriales

Los pesos factoriales nos indican como cada una de las variables iniciales x_i , $i = 1, 2, 3, \dots, 30$ dependen de los factores comunes, es decir que los valores altos en cada componente según la variable indicaran si está da información relevante al factor o si en realidad no le aporta significativamente a este. Por ejemplo en la primer componente las variables más importantes y de mayor aporte son la pregunta 43, la 45, 47, 48, 49 y 50. En cada componente se debe hacer la respectiva revisión de cuales son las preguntas que le aportan mayor información ya que éstas serán las que lo conformarán, de acuerdo a esto se revisan dichas preguntas de manera contextual y le se asigna un nombre a cada componente que permite identificar que evalúa o representa de acuerdo a las preguntas que lo componen.

Tabla 3-1: Pesos factoriales

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
P2	-0.06	-0.11	0.01	-0.10	0.15	-0.04	0.27	0.14	-0.04	0.15	0.45
P4	-0.07	-0.11	0.02	-0.03	-0.05	0.04	0.36	0.32	-0.01	0.01	0.28
P5	-0.01	0.01	-0.04	0.19	0.05	-0.15	-0.07	-0.15	-0.10	0.49	-0.14
P6	-0.01	0.04	-0.08	0.17	-0.05	0.05	0.05	0.17	0.15	0.49	0.06
P11	0.04	0.03	0.41	0.06	0.05	0.01	0.19	-0.08	0.12	0.06	-0.04
P13	0.03	0.01	-0.01	-0.35	0.15	0.04	0.08	-0.05	-0.32	-0.12	-0.23
P14	0.01	0.03	0.06	0.02	-0.01	-0.02	-0.21	-0.05	0.10	-0.37	0.42
P20	0.04	-0.01	0.09	-0.13	-0.02	0.03	-0.31	0.18	-0.09	0.29	0.34
P21	0.04	0.05	0.03	0.03	-0.01	0.15	-0.30	0.08	-0.30	0.11	0.05
P22	0.07	0.08	0.05	0.22	-0.28	-0.01	0.02	0.38	-0.16	-0.16	-0.11
P23	0.04	0.03	0.39	0.08	0.04	0.00	0.24	-0.13	0.03	0.08	-0.08
P24	0.07	0.10	-0.02	-0.15	0.35	-0.02	-0.10	0.28	0.23	0.09	-0.14
P25	-0.04	0.03	-0.02	0.22	0.01	-0.03	-0.15	0.06	0.54	-0.17	0.01
P28	0.09	0.08	-0.05	-0.18	-0.17	-0.28	0.14	-0.10	-0.03	-0.02	0.19
P33	0.09	0.14	-0.06	0.21	0.31	0.16	0.09	-0.15	-0.20	-0.07	0.18
P34	0.09	0.08	-0.05	-0.10	-0.05	-0.31	-0.10	-0.34	0.15	0.13	0.17
P36	0.09	0.11	-0.01	-0.18	0.24	-0.07	0.01	0.31	0.21	0.01	-0.20
P37	0.08	-0.07	0.18	-0.01	-0.03	0.10	-0.28	-0.02	0.02	0.02	0.12
P38	0.10	0.13	-0.07	0.22	0.35	0.09	0.09	-0.06	-0.09	-0.08	0.23
P39	0.10	0.13	-0.04	0.17	-0.19	0.00	0.07	-0.07	-0.08	0.01	-0.04
P40	0.09	0.15	-0.05	-0.15	-0.23	0.36	0.10	-0.04	0.15	0.07	0.01
P41	0.10	-0.13	0.14	-0.10	0.04	0.12	-0.19	-0.01	0.04	0.01	0.06
P43	0.12	0.09	0.01	-0.11	-0.04	-0.41	0.04	0.05	-0.01	-0.00	0.07
P44	0.10	0.14	-0.07	-0.18	-0.20	0.32	0.14	-0.05	0.16	0.10	0.12
P45	0.12	-0.20	-0.06	0.10	0.03	-0.04	0.08	-0.00	0.08	-0.01	0.01
P47	0.11	0.07	0.08	0.11	-0.09	-0.18	-0.06	0.34	-0.21	-0.06	0.01
P48	0.15	-0.15	0.09	-0.06	-0.01	0.10	-0.05	-0.05	0.01	0.04	-0.11
P49	0.14	-0.16	-0.23	0.05	-0.03	0.03	0.03	0.01	-0.01	-0.02	-0.01
P50	0.17	-0.20	-0.08	0.04	0.01	0.05	0.03	-0.03	0.04	0.02	-0.03
P51	0.18	-0.09	-0.03	0.09	0.03	-0.05	0.15	0.01	0.06	-0.05	-0.07

3.2.4. Valores Propios

Los valores propios cuentan como uno de los métodos para determinar el número de factores a conservar. La regla de Kaiser dice que se deben tomar el número de valores propios superiores a la unidad. En este caso si observamos los primeros 11 valores propios, estos se encuentran por encima de la unidad, de esta manera existe un soporte estadístico sobre lo que se decidió teniendo en cuenta la inercia recogida en cada uno.

$$\begin{pmatrix} 3,9575988 \\ 2,9747403 \\ 2,0037423 \\ 1,5262806 \\ 1,4089178 \\ 1,3485833 \\ 1,2454749 \\ 1,1927284 \\ 1,1595712 \\ 1,1243309 \\ 1,0590911 \\ 0,9947755 \\ 0,9703609 \\ 0,8950545 \end{pmatrix}$$

Como se puede observar a partir de valor propio 11 no se tienen valores propios mayores a 1, por lo tanto no se toman más factores de los sugeridos.

3.2.5. Componentes

Sabiendo que el número de componentes a tener en cuenta es 11, cabe resaltar que el método de análisis factorial trabaja con variables estandarizadas las cuales están no correlacionadas

entre sí y tienen $\mu = 0$ y $var = 1$ por lo tanto, los pesos factoriales son las correlaciones entre las variables y los factores creados. Con esta información se pueden crear las combinaciones lineales de cada factor lo cual representa una componente la cual tendrá un nombre propio dependiendo de lo que represente.

En este caso, luego de tener las componentes con su grupo de preguntas se hizo un respectivo debate con la docente especialista en estudios sobre prácticas sexuales de estudiantes y sus comportamientos quien participó en la selección de las 30 variables que harían parte del índice. La especialista sugirió los nombres de cada componente teniendo en cuenta las variables que la conforman y la asociación de las mismas con el mayor o menor riesgo de de contraer ITS y VIH/SIDA.

- **Componente 1=** "Uso del condón "
Compuesta por las variables: P27-P28-P29-P30
- **Componente 2=** "Prácticas sexuales"
Compuesta por las variables: P15-P21-P5-P24
- **Componente 3=** "Vida sexual activa"
Compuesta por las variables: P5-P11-P18-P22
- **Componente 4=** "Trabajo sexual"
Compuesta por las variables: P10-P13-P19
- **Componente 5=** "Promiscuidad"
Compuesta por las variables: P6-P12-P17-
- **Componente 6=** "Bienestar sexual"
Compuesta por las variables: P9-P21
- **Componente 7=** "Tiempo libre"
Compuesta por las variables: P1-P2
- **Componente 8=** "Educación sexual"
Compuesta por las variables: P4-P8-P26
- **Componente 9=** "Diversión"
Compuesta por las variables: P7-P16
- **Componente 10=** "Rechazo sexual"
Compuesta por las variables: P3
- **Componente 11=** "Exposición a contagio"
Compuesta por las variables: P16-P19-P24

De acuerdo a las variables definidas anteriormente se puede observar que representa cada una dentro de la componente.

3.3. Indicador

Teniendo en cuenta las anteriores componentes generadas de manera individual, se plantea la construcción del índice el cual considera los ítems seleccionados del cuestionario inicial y sus características.

Considerando las 30 preguntas a elección de las psicólogas a partir del cuestionario inicial, se segmentan los X_i ítems de la siguiente manera:

- **Respuesta binaria** Los primeros 24 X_i distribuyen Bernoulli con parámetro p , donde $0 \leq p \leq 1$ para todo $i = 1, 2, \dots, 17$.

$$X_i = \begin{cases} 1 & \text{Si cumple con esa característica en sus acciones.} \\ 0 & \text{No es necesaria esa característica en sus acciones.} \end{cases}$$

Respuesta escala Likert Los últimos seis X_i tienen respuesta de escala Likert donde no todos presentan las mismas opciones de respuesta. A continuación se presenta de manera individual las escalas de respuesta.

1.

$$X_i = \begin{cases} 0 & \text{Inactivo sexualmente} \\ 1 & \text{Menos activo sexualmente} \\ 2 & \text{Activo sexualmente} \end{cases}$$

donde $i = 18$

2.

$$X_i = \begin{cases} 1 & \text{Nunca} \\ 2 & \text{Algunas veces} \\ 3 & \text{Casi siempre} \\ 4 & \text{Siempre} \\ 5 & \text{No responde} \end{cases}$$

donde $i = 19, \dots, 25$

3.

$$X_i = \begin{cases} 1 & \text{Nunca ha tenido prácticas sexuales penetrativas} \\ 2 & \text{No ha prácticas sexuales penetrativas en el último año} \\ 3 & \text{Ha tenido prácticas sexuales pero no penetrativas en el último año} \\ 4 & \text{Ha tenido prácticas sexuales penetrativas pero no en el último año} \\ 5 & \text{Ha tenido prácticas sexuales penetrativas en el último año} \end{cases}$$

donde $i = 26$

4.

$$X_i = \begin{cases} 0 & \text{No hay riesgo} \\ 1 & \text{Riesgo mínimo} \\ 2 & \text{Riesgo alto} \\ 3 & \text{Riesgo total} \end{cases}$$

donde $i = 27, \dots, 30$

El nivel de riesgo de contraer ITS-VIH/SIDA (como variable latente) que tiene un estudiante en la población, puede ser modelado utilizando una variable Y , expresable en forma continua con una escala de intervalo. Y es la expresión de riesgo que se comporta de manera aleatoria en la persona, la cual puede tomar valores definidos entre los posibles Y_{1k} , se le puede asignar, a su comportamiento natural, una distribución de probabilidad $A_{Y_{1k}}(y_{1k})$ desconocida, que a su vez está indexada por un conjunto de parámetros β también desconocidos, pero susceptible de ser estimado mediante observaciones de una muestra.

Sea Y la expresión de riesgo, donde Y puede tomar valores definidos entre $[Y_{1k}, Y_{2k}]$:

- Y_{11} a Y_{21} , donde $Y_{11} = 0$ - $Y_{21} = 1$
- Y_{12} a Y_{22} , donde $Y_{12} = 8,55$ - $Y_{22} = 14,80$
- Y_{13} a Y_{23} , donde $Y_{12} = 1,71$ - $Y_{22} = 2,96$

3.3.1. Puntuación por componente

Ya que el método trabaja con los datos estandarizados, para obtener las puntuaciones de cada estudiante se procede a calcular la matriz x estandarizada la cual se estima de la siguiente forma:

$$\frac{(X - \bar{X})/\sigma}{\sqrt{n/n - 1}}$$

Para cada valor de X se debe realizar este procedimiento con esto se obtiene la matriz escalada x de dimensiones (760x30). Para realizar este procedimiento en R se utilizó el comando `xz=scale(xx)*sqrt(760/759)`, donde xx es mi matriz original. Con esta información se procede a evaluar cada estudiante por componente y con esto se generan vectores de $[760, 1]$ por componente. Ahora se utilizarán las 11 componentes para formar una sola matriz de puntuaciones " P " estandarizada la cual quedará de tamaño (760x11).

Con esto se puede decir que:

- Las 11 componentes son variables latentes halladas a partir de 30 variables observadas y éstas a su vez al ser evaluadas pueden ser representadas de la siguiente manera.

$$C_{i,j} = \sum_1^q \lambda_{qj} * x_{i,qj} \quad (3-2)$$

donde:

- λ peso factorial asignado al ítem q , en la componente j .
 - $i=1, \dots, 760$ número de estudiantes que respondió la totalidad cuestionario.
 - $j=1, \dots, 11$ número de componentes generadas a partir del método factorial.
 - q ítems que conforman cada componente. q varia dependiendo de la componente ya que no todas las componentes tienen igual número de ítems.
- La matriz de las puntuaciones para cada individuo esta conformada por las puntuaciones de los individuos en cada componente generando.

$$P = \begin{pmatrix} I1 & C1 & \dots & C11 \\ I1 & C_{1,1} = \sum_1^{q_1} \lambda_{q_1} * x_{1,q_1} & \dots & C_{1,11} = \sum_{a=1}^{q_{11}} \lambda_{q_{11}} * x_{1,q_{11}} \\ \vdots & \dots & \dots & \vdots \\ I760 & C_{760,1} = \sum_1^{q_1} \lambda_{q_1} * x_{760,q_1} & \dots & C_{760,11} = \sum_{a=1}^{q_{11}} \lambda_{q_{11}} * x_{760,q_{11}} \end{pmatrix}$$

Con lo anterior se pueden definir las componentes obtenidas las cuales son combinaciones lineales de los ítems que la conforman y a su vez generan un valor único de acuerdo a la matriz estandarizada obtenida y al peso de cada ítem dentro del factor.

$$c1_i = 0,15 * x_{i,48} + 0,14 * x_{i,49} + 0,17 * x_{i,50} + 0,18 * x_{i,51}$$

$$c2_i = 0,14 * x_{i,33} + 0,15 * x_{i,40} + 0,14 * x_{i,44}$$

$$c3_i = 0,41 * x_{i,11} + 0,39 * x_{i,23} + 0,18 * x_{i,37} + 0,14 * x_{i,41}$$

$$c4_i = 0,22 * x_{i,22} + 0,23 * x_{i,25} + 0,22 * x_{i,38}$$

$$c5_i = 0,35 * x_{i,24} + 0,24 * x_{i,36} + 0,16 * x_{i,13}$$

$$c6_i = 0,14 * x_{i,21} + 0,36 * x_{i,40}$$

$$c7_i = 0,27 * x_{i,2} + 0,36 * x_{i,4}$$

$$c8_i = 0,15 * x_{i,6} + 0,18 * x_{i,20} + 0,34 * x_{i,47}$$

$$c9_i = 0,1 * x_{i,14} + 0,15 * x_{i,34}$$

$$c10_i = 0,5 * x_{i,5}$$

$$c11_i = 0,085 * x_{i,34} + 0,099 * x_{i,38} + 0,99 * x_{i,44}$$

Teniendo en cuenta que $i=1, \dots, 760$ representa el número de estudiantes que participó en el estudio.

Si ejecutamos las anteriores ecuaciones que son combinaciones lineales podemos obtener una puntuación para cada estudiante según su respuesta en esa pregunta, por ejemplo $0,35 * x_{75,22}$ está sería la puntuación del estudiante 75 en la pregunta 22 y así sucesivamente. Como cada componente me arroja un resultado por estudiante para conformar el indicador se tendrá en cuenta la puntuación global por estudiante la cual sera presentada en la próxima sección.

Ahora se construirá el indicador final el cual sale de la suma de los valores obtenidos por cada estudiante en las 11 componentes. Es decir que al final se tiene una columna con 760 valores los cuales representan la puntuación final de riesgo que cada estudiante tiene según sus respuestas.

Con esto se obtiene un único valor para cada estudiante, el cual tendrá en cuenta todas las componentes que representan características de comportamiento y hábitos sexuales las cuales los pueden poner en un mayor o menor riesgo. Ya que el índice se encuentra en un rango de $[-6, 10]$ y, está escala no es práctica al momento de ser puntuado un estudiante porque no va a ser fácil en el ámbito medico y psicológico explicar que estar en -6 es correr mayor riesgo a contraer una infección de transmisión sexual a estar en una puntuación de 10 ya que la escala no es exacta ni su rango se presta para ser interpretado fácilmente porque no van a saber que es mayor o que es menor en el grado de riesgo mientras que, si se coloca todo a una escala uniforme donde cero sea la puntuación mínima y 1 la máxima es mucho mas cómodo de explicar y de interpretar para el experto en salud y el mismo estudiante. Por lo anterior se plantea realizar una transformación la cual consta de restarle a cada valor observado el valor mínimo(-6) y dividirlo sobre el rango(max-min).

$$y_i = \frac{IR[i]_{observado} - IR_{min}}{IR_{max} - IR_{min}}$$

El rango en el que oscilan las puntuaciones finales de los estudiantes corresponde al intervalo $[0, 1]$. Finalmente se obtiene un vector de tamaño Y_{760*1} .

3.4. Otros rangos para el indicador

Ya que se llegó a un posible indicador teniendo en cuenta el método factorial y las ponderaciones de este, se planteó evaluar el indicador con dos nuevas matrices y establecer dos escalas más de puntuación.

3.4.1. Datos originales

Bajo la misma estructura del índice IR creado y presentado en la sección anterior se propone puntuar a los estudiantes pero utilizando la matriz X original, sin ser escalada como la matriz con la que trabaja el análisis factorial. Bajo este caso el índice IR puntúa a los estudiantes entre $[8.55, 14.80]$.

$$IRXi = \sum_{j=1}^{11} X_{i,j}, \quad i = 1, \dots, 760 ; j = 1, \dots, 11$$

donde X es la matriz original de respuestas las cuales están en escala likert y dicotómica.

3.4.2. Normalización de datos observados

Finalmente bajo la misma propuesta de indicador IR se propone evaluar las puntuaciones de los estudiantes aplicando a la matriz X la normalización característica y de esta manera convertirla a una escala uniforme.

$$IRX'i = \sum_{j=1}^{11} X'_{i,j}, \quad i = 1, \dots, 760 ; j = 1, \dots, 11 \quad (3-4)$$

donde X' es la matriz de respuestas originales normalizada. Utilizando la matriz X' en la ecuación 3-4 se obtienen las correspondientes puntuaciones de riesgo para cada estudiante las cuales oscilan en el intervalo $[1.71, 2.96]$.

3.5. Escalas de riesgo

Luego de tener tres formas de escalamiento para el indicador de riesgo se creó dentro de cada rango de puntuación niveles de riesgo basados en los cuartiles de cada rango de datos. Las escalas de riesgo se dividieron de la siguiente manera: el 25 % o menos donde se puedan clasificar los estudiantes se considera riesgo bajo, hasta el 50 % donde se encuentren los estudiantes riesgo moderado y el 75 % o más donde se encuentren los estudiantes como riesgo alto.

Observando la figura **3-2** que representa la distribución del primer índice el cual utiliza la matriz estandarizada del método factorial y, al ser transformadas sus puntuaciones se establece una escala de riesgo entre $[0,1]$ la cual se divide entre:

1. **Riesgo bajo:** Se considera que un estudiante presenta riesgo bajo cuando su puntuación oscila entre $[0,0.2)$.
2. **Riesgo moderado:** Se considera a un estudiante en riesgo moderado cuando su puntuación se encuentra entre $[0.2,0.4)$
3. **Riesgo alto:** Los estudiantes con puntuaciones entre $[0.4,1]$ son estudiantes que deben ser intervenidos ya que se considera que estos están en un alto riesgo de adquirir ITS y VIH/SIDA.

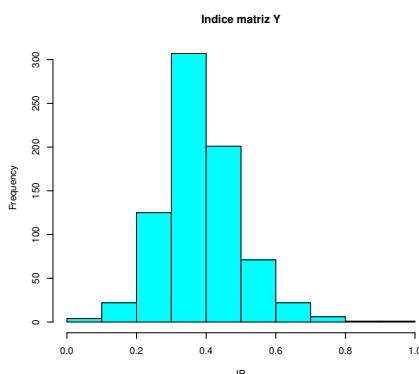


Figura 3-2: Índice IR- Matriz x

El segundo índice se ve representado en la figura **3-3** para el cual se debe tener en cuenta que, este se evaluó utilizando la matriz de respuestas originales de los estudiantes y se generaron puntuaciones en una escala de riesgo en un intervalo de $[8.55,14.80]$ la cual se divide entre:

1. **Riesgo bajo:** Se considera que un estudiante presenta riesgo bajo cuando su puntuación oscila entre $[8.55,10.5)$.
2. **Riesgo moderado:** Se considera a un estudiante en riesgo moderado cuando su puntuación se encuentra entre $[10.55,12)$.
3. **Riesgo alto:** Los estudiantes con puntuaciones entre $[12,14.80]$ son estudiantes que deben ser intervenidos ya que se considera que estos están en un alto riesgo de adquirir ITS y VIH/SIDA.

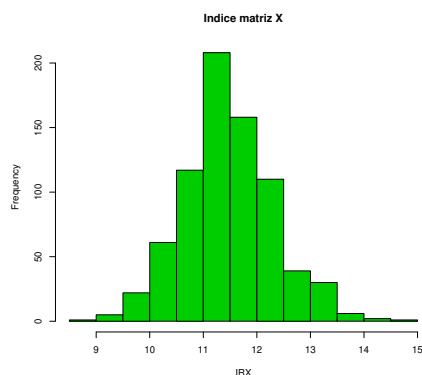


Figura 3-3: Indice IR- Matriz X

La última propuesta para el indicador de riesgo consideró utilizar la matriz normalizada de los datos originales y se puede observar su distribución en la figura **3-4**. En el tercer indicador se establece la escala de riesgo más pequeña de las tres propuestas que oscila en un intervalo de $[1.71, 2.96]$ la cual se divide en:

1. **Riesgo bajo:** Se considera que un estudiante presenta riesgo bajo cuando su puntuación oscila entre $[1.71, 2.1)$.
2. **Riesgo moderado:** Se considera a un estudiante en riesgo moderado cuando su puntuación se encuentra entre $[2.1, 2.3)$
3. **Riesgo alto:** Los estudiantes con puntuaciones entre $[2.3, 2.96]$ son estudiantes que deben ser intervenidos ya que se considera que estos están en un alto riesgo de adquirir ITS y VIH/SIDA.

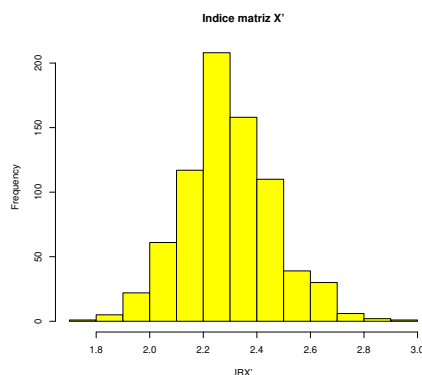


Figura 3-4: Indice IR- Matriz X'

3.6. Puntos de corte

Luego de encontrar una forma para el indicador la cual reúne las características según los comportamientos y hábitos de los estudiantes respecto a sus conductas sexuales, se plantea utilizar una curva ROC No paramétrica, la cual permita encontrar un punto de corte que permite distinguir a partir de que valor según el puntaje que obtenga un estudiante en el *IR* está en bajo, medio o alto riesgo de contraer una ITS - VIH/SIDA. Este punto de corte será global para los estudiantes.

Para construir la curva ROC es necesario contar con un grupo de individuos que puedan ser considerados en riesgo de contraer ITS-VIH/SIDA y otro grupo que se pueda considerar libre de riesgo. El grupo libre de riesgo que se tomó en este estudio fue caracterizado por la pregunta 81 del cuestionario la cual hace referencia al uso persistente del condón, esto quiere decir que se tomó el grupo de estudiantes que utilizaban "*SIEMPRE*" el preservativo en sus relaciones sexuales independiente de pareja(estable o no estable), si planificaba o no, entre otros factores que afectan el uso persistente del condón.

Pregunta 81= ¿Con que frecuencia usas el preservativo u otros métodos de barrera en las prácticas sexuales?

Los estudiantes que contestaron a esta pregunta que siempre y casi siempre los usan se clasificaron como estudiantes que si hacen uso persistente del condón, mientras que los que lo usan a veces, casi nunca y nunca fueron considerados como estudiantes que no hacen uso persistente del condón. Con esta información y las puntuaciones generadas a partir del indicador de riesgo construido se procede a construir la curva ROC, la cual ayudará a definir que tan lejos o cerca están los estudiantes de ser considerados como alto riesgo de contraer ITS y VIH/SIDA.

Sobre el nivel de riesgo que tiene cada estudiante y, teniendo en cuenta las respuestas de los 30 ítems se exponen los niveles de riesgo según los puntos de corte establecidos.

Tabla 3-2: Resumen niveles de riesgo, según matriz de respuestas utilizadas

		Y_1	Y_2
Aproximación	$Y_{1k} - Y_{2k}$	Corte 1	Corte 2
1	0 – 1	0.2	0.4
2	8,55 – 14,80	10.5	12
3	1,71 – 2,96	2.1	2.3

Se construyeron tres curvas ROC con las puntuaciones obtenidas al evaluar las tres propuestas de índices y aplicados a los dos grupos formados a partir de operacionalización efectuada sobre las respuestas de los estudiantes a la pregunta 81 del cuestionario. Se definió cual de los índices logra clasificar con mayor exactitud a los estudiantes en riesgo de contraer ITS y VIH/SIDA dependiendo de la sensibilidad y la especificidad estimada. Teniendo las

tres curvas ROC que son la representación del estado de riesgo en el que se encuentran los estudiantes y permiten establecer según la sensibilidad y especificidad calculada cuantos están bien clasificados en riesgo de contraer ITS y VIH/SIDA ó cuantos de estos no están en riesgo pero son clasificados en riesgo, se estable el punto de corte el cual determinará donde es óptimo el valor de la sensibilidad y especificidad. De está manera el punto de corte representará el límite entre los estudiantes en riesgo reales y los estudiantes en no riesgo reales.

4 Resultados

4.1. Análisis descriptivo

Debido a que el tipo de variables con las que se trabajó en el estudio son categóricas se presentan recuentos de algunas de estas y lo que se evidencia de manera descriptiva al ser cruzadas entre ellas.

Se observó con que frecuencia usan los estudiantes condón al momento de tener una relación sexual a lo que 237 estudiantes respondieron como que siempre o la mayoría de veces lo usaban y 523 estudiantes dijeron no usar condón de manera constante. Por otra parte de los 760 estudiantes se encontró que 705 de estos nunca ha tenido prácticas sexuales con más de una persona a la vez, 50 estudiantes en algunos casos tuvieron prácticas sexuales con varias personas al mismo tiempo y 5 estudiantes dijeron que casi siempre sostienen prácticas sexuales con varias personas al mismo tiempo. Al momento de preguntarles a los estudiantes por la percepción que estos tienen sobre diversión y la relación de esta con el hecho de ser relevante tener sexo siempre que salen a divertirse se encontró que 670 estudiantes el 88 % de la muestra consideran la necesidad de tener sexo en sus salidas, mientras que 90 estudiantes dijeron que no era necesario tener sexo para divertirse.

Según algunos estudios las personas que tienen una relación formal o pareja estable disminuyen el uso del preservativo ya que, depositan un alto nivel de confianza en esta. Teniendo en cuenta lo anterior, se analizó las preguntas 8 y 20 las cuales representan información sobre el uso del condón de acuerdo al tipo de relación y el concepto de diversión relacionado con sexo. En la tabla 4-1 se puede observar los resultados al cruzar las dos preguntas donde se evidencia que 568 estudiantes que tienen una relación formal usan condón y necesitan tener relaciones sexuales para sentirse bien cuando salen a divertirse. De los 192 estudiantes restantes, 102 dejan de usar condón al tener una pareja estable y, además de eso siempre que salen tienen una relación sexual la cual no necesariamente es con su pareja. Por otra parte 73 estudiantes dijeron usar condón en sus relaciones formales pero no sienten que es necesario tener relaciones sexuales durante sus actividades de ocio. Finalmente un grupo de 17 estudiantes expresaron no usar condón cuando tienen una pareja estable pero tampoco necesitan tener una relación sexual siempre que salen a divertirse.

Usa condón en las relaciones formales	Siempre que sale tiene sexo		
	Si	No	Total
SI	568	73	641
NO	102	17	119
Total	670	90	760

Tabla 4-1: Uso del preservativo vs Tener sexo siempre al salir

Utilizando la pregunta 30 que representa la creación de la variable denominada "Riesgo 1" la cual, según los criterios clínicos utilizados por las enfermeras que tuvieron en cuenta la relación que existe al recibir alguna retribución económica o material a cambio de sexo y, el uso del preservativo en el último año en sus prácticas sexuales lo cual se obtiene de las preguntas 64 y 70 del cuestionario original, determinaron una escala de riesgo teórica que, clasificaba a los estudiantes en no riesgo, riesgo mínimo, riesgo alto y riesgo total. Dicha clasificación depende de la manera en que estos respondan a las preguntas 64 y 70 y se puede observar en la tabla 4-2. Al revisar el cruce de estas dos preguntas se encontró que de los 760 estudiantes a los que se les aplicó el cuestionario 720 habían recibido retribución económica o material y, 234 de estos presentaba un riesgo algo por el uso inadecuado del preservativo durante el último año mientras que 4 se encuentran en riesgo total. De las 40 personas que no han recibido retribución alguna 29 fueron clasificadas en un alto riesgo y solo 5 no presentan riesgo, bajo el criterio creado por las enfermeras.

Retribución económica o material por sexo	Uso de condón en su última relación sexual				Total
	No riesgo	Riesgo mínimo	Riesgo alto	Riesgo total	
SI	66	416	234	4	720
NO	5	6	29	0	40
Total	71	422	263	4	760

Tabla 4-2: Retribución económica o material a cambio de sexo y, Uso del condón.

A manera general el 92 % de los estudiantes encuestados no ha tenido ITS y el 8 % de si tuvo ITS en algún momento de su vida. De los estudiantes que tuvieron ITS 46 practicó intercambio de parejas y 17 no participó de dicha práctica sexual. Por otra parte 539 estudiantes practican intercambio de parejas pero no han tenido ITS y 158 que no han tenido ITS no realizan esa práctica.

4.2. Indicador

En el capítulo anterior se creó una sola estructura de indicador basada en el Análisis Factorial pero, se evaluó la propuesta del indicador con tres matrices las cuales contienen las respuestas de los estudiantes en diferentes escalas. La primer matriz consta de las respuestas de los individuos que están escaladas como lo realiza el Análisis Factorial, la segunda matriz de respuestas de los estudiantes con sus respuestas originales y la tercera es una estandarización

a las respuestas originales de los estudiantes. Cabe resaltar que las puntuaciones obtenidas con la matriz escalada del método factorial se les realizó una transformación por motivos de comodidad y facilidad al momento de ser interpretadas por las enfermeras.

Normalidad y distribución

A pesar de que se contó con un tamaño de muestra grande (760 estudiantes) esto no daba seguridad absoluta de que el comportamiento o la manera en que distribuyen las puntuaciones del indicador se ajustara a una distribución normal, teniendo en cuenta que en la Sección 3.4.2 se explica la normalización que realiza el análisis factorial a su matriz de datos. Por esta razón, se aplicó un test de normalidad donde se pretende observar para cada indicador como es su comportamiento teniendo en cuenta que la variación de cada uno es la matriz de respuestas con la que se evalúa.

$$H_0 = Y_i \sim N(0, 1)$$

$$H_0 = Y_i \approx N(0, 1)$$

1. Para el primer indicador no se aplicó prueba de normalidad ya que en este se utilizó la matriz X normalizada con la que trabaja el análisis factorial.
2. Las pruebas de normalidad para el indicador 2 y 3 planteados arrojaron el mismo resultado esto se debe a que es la misma matriz solo que una presenta una transformación de escala.

Se utilizó la prueba de normalidad Kolmogorov Smirnov la cual bajo un $\alpha = 0,05$ arrojó un valor $p - value = 0,02731$ y, teniendo en cuenta la hipótesis de normalidad:

Se rechaza la hipótesis nula, por lo tanto, no se puede afirmar que la distribución de las puntuaciones de los estudiantes para el indicador de la propuesta 2 y 3 se ajustan a un modelo de probabilidad normal, según el test de Kolmogorov Smirnov.

4.2.1. Curva ROC-Punto de corte

La curva ROC se utilizó como herramienta para clasificar a los estudiantes lo mas correctamente posible en estado de riesgo y no riesgo de contraer ITS-VIH/SIDA. Para poder clasificarlos fue necesario establecer un punto de corte donde se consideren en "riesgo" aquellos estudiantes con puntuaciones mayores o iguales al punto de referencia y, en "no riesgo" aquellos estudiantes con puntuaciones menores a dicho valor. El valor de referencia es establecido observando la sensibilidad y especificidad del test.

Otra de las razones por las cuales se emplea en el desarrollo del trabajo la curva ROC, es la falta de normalidad en la distribución de las puntuaciones de los estudiantes, esto da indicios de utilizar curvas ROC-No paramétricas que ayuden a obtener el punto de corte o valor de referencia.

Antes de presentar el punto de corte en cada índice que permite clasificar a los estudiantes en "riesgo" (Aquellos estudiantes que NO siempre usan preservativo) y "no riesgo" (Aquellos estudiantes que siempre usan preservativo), se expone la escala de cada índice y, según la pregunta 81 del cuestionario (Pregunta tomada como el patrón de oro) donde se observa a los estudiantes que siempre usan preservativo en sus relaciones sexuales, se expone la sensibilidad y especificidad calculada para cada indicador de riesgo.

1. **Índice 1** Teniendo en cuenta lo anterior para las puntuaciones de los Y_i los estudiantes que se encuentren por encima del 84 % de la sensibilidad se clasificarán como jóvenes en "riesgo" y los que estén por debajo serán individuos en "no riesgo" de contraer ITS y VIH/SIDA.

$$T_1 = \begin{cases} 1 & \text{si } Y \geq 0.5644 \\ 0 & \text{si } Y < 0.5644 \end{cases}$$

donde obtener una puntuación de 0.5644 es el valor que representa el 84 % de la sensibilidad en la escala del indicador 1. A continuación se presenta en la figura 4-1 la curva ROC para el índice 1, la cual presenta un AUC=0,26.

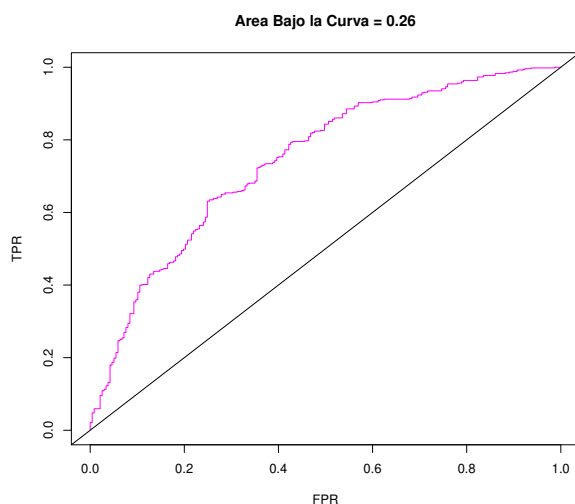


Figura 4-1: Curva ROC Indicador 1

Observando la tabla 5-1 la cual según el cruce entre la pregunta 81 sobre los que hacen uso persistente de condón y los estudiantes clasificados como "riesgo" ó "no riesgo" y, de acuerdo a la curva ROC 4-1 se puede determinar que de los 128 estudiantes clasificados

en "riesgo(1)", el 78 % están correctamente clasificados y el 22 % son estudiantes mal clasificados en "riesgo" de contraer ITS-VIH/SIDA ya que estos si hacen uso persistente del condón. Por otra parte 632 estudiantes fueron clasificados como estudiantes en "no riesgo(0)" de los cuales el 67 % fue clasificado de manera incorrecta al no hacer uso constante del preservativo y solo el 33 % de los estudiantes fueron bien clasificados como en no riesgo ya que hacen uso constante del preservativo.

	Uso de condón		Total
	No uso condón	Uso de condón	
1	100	28	128
0	423	209	632
Total	523	237	760

Tabla 4-3: Clasificación Indicador 1

Con esta información y, luego de establecer el punto de referencia con ayuda de la curva ROC. Se expone la sensibilidad y especificidad del primer indice ya que se tiene la cantidad de falsos positivos y negativos por parte del primer indice.

■ **SENSIBILIDAD(Correctos clasificados en riesgo Indice 1):**

$$\frac{100}{523} * 100 = 19 \%$$

Con un error estándar de sensibilidad diagnostica de 1.7 % y, un intervalo de confianza al 95 % entre [16 %, 23 %]. También se evaluó un intervalo de confianza al 99 % entre [15 %, 24 %]

■ **ESPECIFICIDAD(Correctos clasificados en no riesgo Indice 1):**

$$\frac{209}{237} * 100 = 88 \%$$

Con un error estándar de especificidad diagnostica del 8 % y, un intervalo de confianza al 95 % entre [72 %, 100 %]. De igual manera se evaluó un intervalo de confianza al 99 % entre [67 %, 100 %]

2. **Indice 2 y 3** Antes de presentar los resultados para el indice (2 y 3) se recuerda que los resultados obtenidos para estas dos propuestas son exactamente iguales ya que las matrices evaluadas tienen una diferencia de escala la cual no afecta los resultados finales ni las conclusiones. Teniendo en cuenta lo anterior para las puntuaciones de los X_i y x'_i se clasificarán los individuos como estudiantes en riesgo y estudiantes en no riesgo. Los estudiantes que se encuentren por encima del 84 % estarán en riesgo y los que estén por debajo serán individuos en no riesgo de contraer ITS y VIH/SIDA.

$$T_2 = \begin{cases} 1 & \text{si } Y \geq 12.70 \\ 0 & \text{si } Y < 12.70 \end{cases}$$

$$T_3 = \begin{cases} 1 & \text{si } Y \geq 2.54 \\ 0 & \text{si } Y < 2.54 \end{cases}$$

donde 12.70 y 2.54 es el valor que representa el 84 % de la sensibilidad en la escala el indicador 2 y 3 respectivamente. A continuación se presenta en la figura 4-2 la curva ROC para el índice 2 y 3 donde se evidencia que el AUC disminuye 0.06 % respecto al del indicador 1.

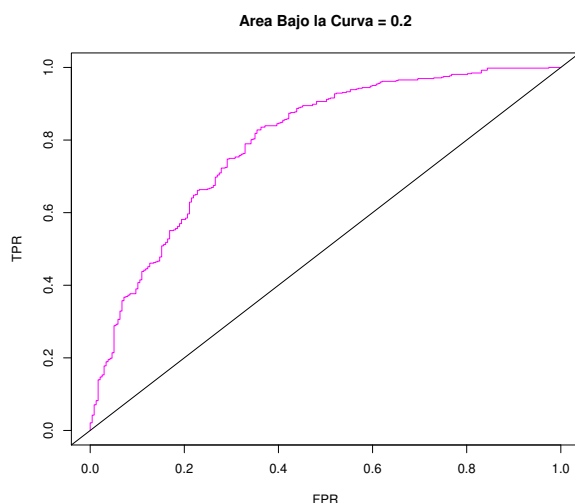


Figura 4-2: Curva ROC Indicador 2 y 3

Observando la tabla 4-4 la cual según el cruce entre la pregunta 81 sobre los que hacen uso persistente de condón y los estudiantes clasificados como "riesgo" ó "no riesgo" y, de acuerdo a la curva ROC 4-2 se puede determinar que de los 110 estudiantes clasificados en "riesgo(1)", el 78 % están correctamente clasificados y el 22 % son estudiantes mal clasificados como en "riesgo" de contraer ITS-VIH/SIDA ya que estos si hacen uso persistente del condón. Por otra parte 650 estudiantes fueron clasificados como estudiantes en "no riesgo(0)" de los cuales el 67 % fue clasificado de manera incorrecta al no hacer uso constante del preservativo y solo el 33 % de los estudiantes tuvieron una correcta clasificación como en no riesgo ya que hacen uso constante del preservativo.

	Uso de condón		
	No uso condón	Uso de condón	
1	86	24	110
0	437	213	650
Total	523	237	760

Tabla 4-4: Clasificación Indicador 2 y 3

Con esta información y, luego de establecer el punto de referencia con ayuda de la curva ROC. Se expone la sensibilidad y especificidad del primer índice ya que se tiene la cantidad de falsos positivos y negativos por parte del segundo y tercer índice.

■ **SENSIBILIDAD(Correctos clasificados en riesgo Índice 2 y 3):**

$$\frac{86}{523} * 100 = 16 \%$$

Con un error estándar de sensibilidad diagnóstica de 1.6 % y, un intervalo de confianza al 95 % entre [13 %, 20 %]. También se evaluó un intervalo de confianza al 99 % entre [12 %, 21 %]

■ **ESPECIFICIDAD(Correctos clasificados en no riesgo Índice 1):**

$$\frac{213}{237} * 100 = 90 \%$$

Con un error estándar de especificidad diagnóstica del 8 % y, un intervalo de confianza al 95 % entre [74 %, 100 %]. De igual manera se evaluó un intervalo de confianza al 99 % entre [68 %, 100 %]

A pesar de que las diferencias numéricas entre las propuestas de los índices no son grandes en términos de sensibilidad y especificidad se prefiere utilizar la propuesta del indicador 1 ya que, clasifica 3 % mejor a los estudiantes en riesgo de contraer ITS y VIH/SIDA.

5 Uso del Indicador de Riesgo

A continuación se presentan los resultados obtenidos en un estudio de simulación que se realizó de 1000 respuestas de estudiantes teniendo en cuenta un entorno controlado el cual ayuda a minimizar la variación. Por otra parte se expone los resultados obtenidos al aplicar los índices por universidad.

5.1. Estudio de Simulación

Se realizó un estudio de simulación con el objetivo de analizar la eficacia del índice obtenido y la respuesta que este generaba al evaluarlo con la información recolectada de los estudiantes. Además de que la simulación permite observar como sería su rendimiento al ser expuesto a respuestas de estudiantes en un contexto controlado y con las respectivas mejoras en el cuestionario.

Para esto, se realizó una simulación de 1000 respuestas de estudiantes con ayuda del software "R" el cual simulaba para cada estudiante una respuesta en determinada pregunta teniendo en cuenta la escala y las opciones a esta es decir, si era escala likert en esa pregunta se simulaba una respuesta entre (1-5), si era dicotómica se simulaba una respuesta entre (0,1) y así sucesivamente en cada pregunta por cada estudiante simulado. A estas respuestas se las evaluó bajo el indicador 1 ya que este fue el mejor.

Teniendo en cuenta lo anterior se procede a observar en la figura ??ón Índice simuladoDistribución Índice simuladoistribución que presenta las puntuaciones del índice bajo la matriz de respuestas simuladas en un escenario inverso a lo que fue la realidad. Las puntuaciones generadas por este índice son gráficamente muy simétricas lo cual puede ser consecuencia de la escala a la que este se convierte y al entorno controlado bajo el cual se simula.

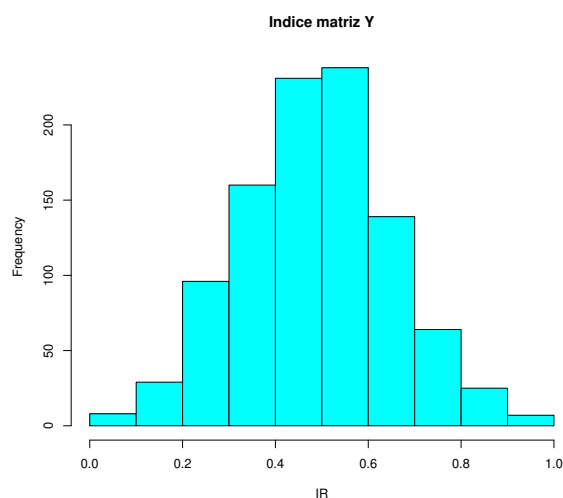


Figura 5-1: Distribución Índice 1 con información simulada

Ahora se observa la curva ROC **5-1** la cual se construye con la matriz de respuestas "Y" simuladas. Comparando la exactitud con la que la curva ROC identifica estudiantes en riesgo y no riesgo se encuentra que esta tiene una mejor percepción al momento de clasificarlos, teniendo en cuenta su AUC.

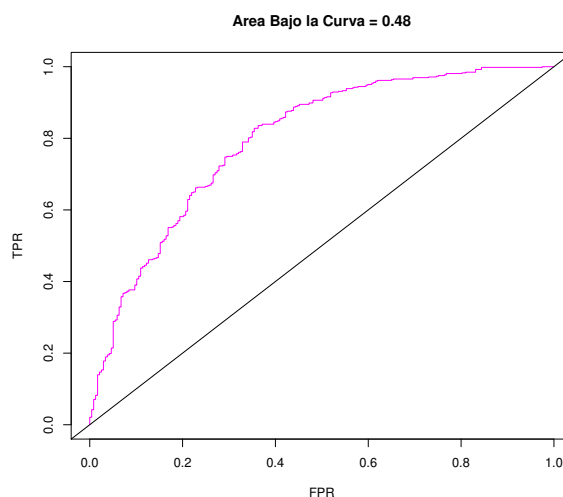


Figura 5-2: Curva ROC Índice 1 simulado

Finalmente se presenta la tabla de falsos positivos y negativos donde se encuentra la discriminación que hace el test teniendo en cuenta la pregunta 81 "uso del condón persiste":

	Uso de condón		Total
	No uso condón	Uso de condón	
1	573	106	679
0	23	298	321
Total	596	404	760

Tabla 5-1: Clasificación Indicador 1 datos simulados

Con esta información se calcula la especificidad y sensibilidad del índice creado bajo la respuesta de los estudiantes en un escenario controlado.

■ **SENSIBILIDAD(Correctos clasificados en riesgo Índice 1 simulado):**

$$\frac{573}{596} * 100 = 96 \%$$

Con un error estándar de sensibilidad diagnóstica de 0.7 % y, un intervalo de confianza al 95 % entre [95 %, 98 %]. También se evaluó un intervalo de confianza al 99 % entre [94 %, 98 %]

■ **ESPECIFICIDAD(Correctos clasificados en no riesgo Índice 1 simulado):**

$$\frac{298}{404} * 100 = 74 \%$$

Con un error estándar de especificidad diagnóstica del 1 % y, un intervalo de confianza al 95 % entre [72 %, 76 %]. De igual manera se evaluó un intervalo de confianza al 99 % entre [71 %, 76 %]

Comparando con los resultados del índice bajo las respuestas reales, este presenta una mejoría del 77 % ya que logra identificar el 96 % de las veces a los estudiantes que se encuentra en riesgo de contraer ITS y VIH/SIDA. Y el 74 % de las veces que se evalúa un estudiante con este índice se clasifica de manera correcta como en no riesgo.

5.2. Aplicación del índice de riesgo por universidad

Luego de obtener las propuestas de indicadores se realiza una aplicación por universidad para determinar el nivel de riesgo en el que estos se encuentran. Teniendo en cuenta el capítulo anterior es posible clasificar cada grupo de estudiantes en un nivel de riesgo con el fin de que sean intervenidos si es necesario y de esta manera evitar contagio de alguna ITS o en caso extremo de VIH. Además de sensibilizar a los estudiantes sobre ITS y VIH, se espera contribuir a las buenas prácticas sexuales las cuales evitan embarazos no deseados el cual es otro problema en los jóvenes universitarios.

5.2.1. Universidad 1

De los 760 estudiantes tenidos en cuenta para la creación del índice por ser aquellos que respondieron en totalidad el cuestionario inicial se encontró al dividirlos por universidad que 163 estudian en la Universidad 1. A estos 163 se les aplicó el índice 1 (se aplica el índice 1 ya que este fue la mejor propuesta respecto a su especificidad y sensibilidad, comparado con los resultados que arrojaba el índice 2) para ser clasificados en un nivel de riesgo de lo cual se encontró que:

- **Índice 1** Según la escala establecida para el índice 1 se tiene que los estudiantes con puntuaciones iguales o superiores a 0,56 (Teniendo en cuenta que el índice 1 puntúa entre $[0, 1]$) están en riesgo y, aquellos que se encuentren por debajo de este valor no presentan riesgo de contraer ITS o VIH. Se encontró los siguientes resultados:
 1. El 73 % de los estudiantes de la Universidad 1 no presentan riesgo de contraer ITS o VIH.
 2. El 27 % de los estudiantes se clasifica en riesgo de contraer ITS o VIH.
 3. Las puntuaciones de los estudiantes en no riesgo oscilan entre $[0,23, 0,42]$.
 4. El puntaje mas alto obtenido para un estudiante de la Universidad 1 fue de 0,88.
- Se evaluó la sensibilidad y especificidad de la clasificación de este grupo de estudiantes donde se encontró que:
 1. **Sensibilidad** Se encontró que el índice 1 logró clasificar de manera correcta como en "RIESGO" el 17,3 % de los estudiantes.
 2. **Especificidad** Por otra parte el índice 1 es mejor clasificando estudiantes en "no riesgo" y, para este grupo de estudiantes se evidenció que el 88 % de las veces clasificó correctamente estudiantes que no estaban en riesgo de contraer ITS y VIH.

5.2.2. Universidad 2

Los estudiantes de la Universidad 2 que respondieron en totalidad al cuestionario inicial fueron 267. A estos se les aplicó la propuesta del índice 1 y se encontró los siguientes resultados.

- **Índice 1** Bajo la misma clasificación anterior sobre que, los estudiantes que arrojen puntuaciones iguales o superiores a 0,56 se clasifican como estudiantes en riesgo se encontró:
 1. El 46.4 % de los estudiantes de la Universidad 2 no presentan riesgo de contraer ITS o VIH. Es decir que más de la mitad de los estudiantes un 53.6 % se encuentra en riesgo de contraer ITS o VIH.

2. Las puntuaciones de los estudiantes en no riesgo oscilan entre $[0,17, 0,52]$.
 3. El puntaje mas alto obtenido para un estudiante de la Universidad 2 fue de 0,96.
- De igual manera se evaluó la sensibilidad y especificidad de la clasificación de este grupo de estudiantes donde se encontró que:
 1. **Sensibilidad** Se encontró que el índice 1 logró clasificar de manera correcta como en "RIESGO" el 18 % de los estudiantes.
 2. **Especificidad** Los estudiantes correctamente clasificados en "no riesgo" son el 93 %.

5.2.3. Universidad 3

De los 760 estudiantes, 219 hacen parte de la Universidad 3. De igual manera se le aplicó a este grupo de estudiantes el índice 1 evidenciando:

- **Índice 1** Teniendo en cuenta las condiciones de clasificación anterior según el puntaje de los estudiantes.
 1. El 77 % de los estudiantes de la Universidad 3 no presentan riesgo de contraer ITS o VIH.
 2. Existe una baja proporción de estudiantes en riesgo del 23 %
 3. Las puntuaciones de los estudiantes en no riesgo oscilan entre $[0,21, 0,48]$.
 4. El puntaje mas alto obtenido para un estudiante de la Universidad 3 fue de 0,92.
- De igual manera se evaluó la sensibilidad y especificidad de la clasificación de este grupo de estudiantes donde se encontró que:
 1. **Sensibilidad** Se encontró que el índice 1 logró clasificar de manera correcta como en "RIESGO" el 22 % de los estudiantes.
 2. **Especificidad** Los estudiantes correctamente clasificados en "no riesgo" son el 86 %.

5.2.4. Universidad 4

Finalmente dentro de los 760 estudiantes existe un pequeño grupo que representa la Universidad 4 de 111 estudiantes.

- **Índice 1** Teniendo en cuenta las condiciones de clasificación anterior según el puntaje de los estudiantes.

1. El 68 % de los estudiantes de Universidad 4 no presentan riesgo de contraer ITS o VIH. Con esto se tiene que existe una proporción de estudiantes en riesgo del 32 %, la cual puede ser alta pero depende del criterio de los expertos.
 2. Las puntuaciones de los estudiantes en no riesgo oscilan entre $[0,18,0,53]$.
 3. El puntaje mas alto obtenido para un estudiante de la Universidad 4 fue de 0,95.
- De igual manera se evaluó la sensibilidad y especificidad de la clasificación de este grupo de estudiantes donde se encontró que:
1. **Sensibilidad** Se encontró que el índice 1 logró clasificar de manera correcta como en "RIESGO" el 21 % de los estudiantes.
 2. **Especificidad** Los estudiantes correctamente clasificados en "no riesgo" son el 76 %.

6 Conclusiones y recomendaciones

6.1. Conclusiones

Estadísticamente el método factorial es una herramienta recursiva y de fácil manejo al momento de crear o medir una variable latente ya que permite agrupar correctamente las características de los individuos y proporcionar un valor equivalente a la realidad de estos así se este trabajando con variables categóricas. El índice desarrollado funcionó mejor al trabajar con la matriz escalada del método factorial ya que logró identificar más individuos en riesgo de contraer ITS y VIH/SIDA que, al ser evaluado por la información original.

Esto también puede ser identificado en la curva ROC y su punto de corte ya que funciona mejor con el primer indicador de riesgo y su AUC a pesar de ser bajo es 0.6 % más alto que en los indicadores 2 y 3 indicando una mejor clasificación a estudiantes en riesgo y no riesgo. Si se da un buen manejo inicial y se identifican las preguntas que más identifiquen riesgo en un estudiante se podría utilizar de nuevo está metodología planteada recalculando los pesos de las componentes y observando si el indicador presenta una mejoría o no.

A diferencia de otras metodologías como el Análisis de correspondencias múltiple el método factorial tiene en cuenta que existen niveles dentro de las respuestas los cuales, pueden dar una mayor importancia al momento de ser analizados generando una aproximación numérica correcta. Por otra parte, el entorno universitario crea ambientes muy cómodos en los que los estudiantes pierden la percepción de responsabilidad sexual y empiezan a aplicar otras prácticas sexuales como recibir dinero a cambio, intercambio de parejas, no usar preservativo sin importar el tipo de pareja, entre otras.

Los índices 2 y 3 arrojaron los mismos resultados ya que lo único que se hizo fue transformar los valores de las respuestas a otra escala, pero estos no sufrieron ninguna variación como la tienen los del primer indicador, los cuales están sujetos al análisis factorial.

Teniendo en cuenta los resultados obtenidos por el índice con los datos simulados en un ambiente mucho más controlado, se observa que este clasifica e identifica correctamente a la mayoría de individuos en riesgo, es decir que se espera que si en algún momento alguien utilice esta metodología sean fiables sus resultados.

6.2. Recomendaciones

Teniendo en cuenta la metodología utilizada para crear el cuestionario y desarrollarlo se sugiere a las personas que deseen iniciar un proyecto igual que no se involucre gran cantidad de personas, se recomienda desarrollar el proyecto con los especialistas necesarios en el tema y desde un inicio contar con la parte estadística ya que con esto se evita correr el riesgo de generar mucho sesgo en lo que se quiere preguntar y se pretende estudiar. Por este motivo no se incluye la evaluación puntual de la validación del test ya que está no va a cumplir por motivos de altas correlaciones en algunas preguntas y no cumplimiento con lo que se esperaba saber.

Es de suma importancia desarrollar este trabajo con el acompañamiento de un especialista en el área de la salud y relaciones estudiantiles esto, debido a que la percepción que se tiene de la sexualidad va muy relacionada con el entorno, religión, vivencias y demás características propias de una persona.

Bibliografía

Alex C. Michalos, e. a. (2014), *Encyclopedia of Quality of life and Well-Being Research*, Springer Science + Bussiness Media Dordrecht.

Barrios, J. (2002), 'Creacion de indices de gestion hospitalaria mediante analisis de componentes principales', *SciElo* **44**(6), 533–540.

Chacon, T. & Corrales, D. (2009), 'Its y sida en adolescentes: descripción, prevención y marco legal', *SciElo* **26**(2), 79–98.

de Naciones Unidas), O. (2000), *ONUSIDA*.

URL: <http://www.unaids.org/es/aboutunaids/unitednationsdeclarationsandgoals/2001declarationofcommit>

de salud OMS, T., ed. (2016), *Factores de riesgo*.

Elkin, C. (2010), Estimacion del indicador de calidad de vida para el departamento de antioquia, Estudios de opinion, Universidad de Antioquia.

Evelyne, V. (2015), 'Clustvarlv: An r package for the clustering of variables around latent variables', *CONTRIBUTED RESEARCH ARTICLES* **7**, 134–148.

Giménez, M.-D. S. (2016), 'Condomes:guía del usuario'.

URL: <https://www.medicina21.com/Especialidad/ginecologia/V683/Condomes-guia-del-usuario.html>

Giron, e. (2013), 'Hiv sexual risk behaviors in youth 15-24 years of age in cali, colombia: Do differences exist among neighborhoods?', *Colombia Médica* **44**(1), 72–79.

Gurrea, M. T. (n.d.), Análisis de componentes principales.

Hand, W. J. K. . D. J. (2009), *Roc Curves for Continuos Data*, number 226, CRC Press.

Hernandez, R. & Lopez, Y. (2008), Importancia del grupo terapéutico de autoapoyo para personas que viven con vih, asi como para personas que conviven con ellos, tesis de licenciatura, facultad de estudios superiores zaragoza, universidad nacional autónoma de mexico, mexico.

José Arturo Granados, C. T. . G. D. (2009), 'The experience of rejection in homosexual university students from mexico city and risk situations for hiv/aids', **51**(6).

- McCutcheon, J. A. H. . A. L. (2002), *Applied Latent Class Analysis*, number 454, United States by Cmabridge University Press, New York.
- Minsalud (2016), *Prevencion*, MINSALUD.
- OMS (2016), Ist(infecciones de transmision sexual), Nota descriptiva 110, OMS(Organizacion Mundial de la Salud).
- OMS (2017a), Technical report.
- OMS (2017b), Vih/sida, Nota descriptiva 360, OMS(Organizacion Mundial de la Salud).
- Online, P. (n.d.), ‘Psicología experimental: Introducción a la psicometría’, *Psicología Online* .
- ONU (1989), *Manual de Indicadores Sociales*, 63 edn, Naciones Unidas. Serie F N° 49.
- Stephane Dray, e. (2017), *ADE4*, 1.7-8 edn, CRAN.R-project.org, <https://cran.r-project.org/web/packages/ade4/ade4.pdf>.
- Teodoro Barros, Dimitri Barreto, e. (2001), ‘Un modelo de prevención primaria de las enfermedades de transmisión sexual y del vih/sida en adolescentes’, *Panam Salud Publica* **10**(2), 86–94.
- Today-CONDOMS (n.d.), Comportamientos de riesgo. tu-sexualidad.
- Tornimbeni, S. (2008), *Introduccion a la psiometria.*, Editorial Paidos SAICF.
- Velásquez, A. (2005), ‘Efecto de un cd multimedia (planeta riesgo xero) en los conocimientos, actitudes y prácticas sobre enfermedades de transmisión sexual y vih/sida de adolescentes de colegios de lima’, *Anales de la Facultad de Medicina* **66**(3).
- William, J. (2015), ‘Homonegative attitudes and risk behaviors for hiv and other sexually transmitted infections among sexually active men in the united states’, *American Journal of Public Health* **105**(12), 2466–2472.
- y Espada, F. (2009), ‘Consumo de sustancias y conductas sexuales de riesgo para la transmisión del vih en una muestra de estudiantes universitarios’, *Anales de psicología* **25**(2), 344–350.