

Facultad de Ciencias

Grado En Estadística

Trabajo De Fin De Grado



**UNIVERSIDAD
DE GRANADA**

Análisis de Modelos de Clase Latente:

Aplicación al caso de malos tratos a las mujeres

Presentado por:

D^a. Sara Pérez-Mallaina de Aramburu

Tutores:

D. Juan de Dios Luna del Castillo

D. Miguel Ángel Montero Alonso

Curso académico:

2023 /2024

Declaración: Declaro que el presente Trabajo de Fin de Grado es original, no habiéndose utilizado fuente sin ser citada debidamente

Autoevaluación

La elección del tema para el Trabajo de Fin de Grado estuvo motivada por dos razones principales. En primer lugar, quería que mis tutores fueran D. Juan de Dios Luna del Castillo y D. Miguel Ángel Montero, ya que les admiro mucho como docentes y como investigadores. En segundo lugar, el tema “violencia de género” me llama mucho la atención y me motiva mucho el investigar y profundizar más en este asunto, por la problemática social en la que nos encontramos.

Aunque tenía la dificultad de que nunca había tratado con Modelos de Clase Latente, decidí elegirlo y aprender sobre ello. Principalmente tuve que ponerme a leer mucho sobre el tema; tanto ensayos, como libros que trataban sobre los Modelos de Clase Latente. Tras las lecturas y familiarizarme con los conceptos de las Clases Latentes, empecé con la limpieza de datos, este comienzo me gustó mucho, ya que se basaba en la asignatura “Diseño de Encuestas” que tanto disfrute en su época. Aunque lo hice con gusto, tardé más de lo debido en prepararlos de manera que el Modelo de Clases Latentes diese sus frutos.

Al llegar a la parte de hacer los modelos latentes fue donde me fui atascando; tuve que probar muchos códigos en R, muchas librerías, hasta que llegué al correcto. Una vez hecho con una variable, el resto de los ítems fueron más agradecidos y sencillos de tratar.

La parte de crear y redactar el documento fue la más sencilla. Para ello también tuve que leer muchos artículos. El hecho de que hubiese tanta información al alcance lo hizo más difícil, ya que había que verificar cada paso y limitarse a dar la información necesaria.

Al finalizar el Trabajo, me he dado cuenta de todos los conocimientos que he adquirido y el esfuerzo ha merecido la pena. En un futuro, espero que sirva para una comprensión sencilla de las características de la violencia de género. Aspiro a que mis hallazgos contribuyan a una mejor comprensión de esta problemática y apoyen el desarrollo de políticas y programas más efectivos para combatir la violencia contra las mujeres.

Resumen

En el presente trabajo se abordará un análisis detallado de los Modelos de Clase Latente (LCA) aplicado al caso de los malos tratos hacia las mujeres. Inicialmente, se proporcionará una explicación exhaustiva sobre lo que constituye la violencia de género, incluyendo la descripción de cada una de las escalas utilizadas y los diferentes ítems que permiten determinar si una mujer sufre o no algún tipo de violencia.

Posteriormente, se analizará la macroencuesta de violencia contra la mujer realizada en España en 2019, detallando minuciosamente cada uno de sus apartados.

Seguidamente, se examinarán en profundidad los Modelos de Clase Latente (LCA), explicando su fundamentación teórica y metodológica, y se aplicarán a los datos obtenidos de la macroencuesta previamente mencionada.

Finalmente, se calcularán las prevalencias tradicionales y las obtenidas a través del Modelo de Clase Latente, con el propósito de compararlas y evaluar la relevancia y ventajas de este último análisis en el contexto de la investigación sobre la violencia de género.

Summary

This paper will present a detailed analysis of Latent Class Analysis (LCA) applied to the case of domestic violence against women. Initially, an exhaustive explanation of what constitutes gender violence will be provided, including descriptions of each scale used and the various items that help determine whether a woman is experiencing any form of violence.

Subsequently, the study will analyze the 2019 macro-survey on violence against women conducted in Spain, detailing each section thoroughly.

Next, Latent Class Analysis (LCA) will be examined in depth, explaining its theoretical and methodological foundations, and it will be applied to the data obtained from the aforementioned macro-survey.

Finally, traditional prevalence rates and those obtained through the Latent Class Model will be calculated, compared, and evaluated to assess the relevance and advantages of this latter analysis in the context of research on gender violence.

ÍNDICE

1. Introducción.....	7
1.1. Qué es y cómo se mide la violencia contra las mujeres (violencia de género).....	8
1.1.1. Definición.....	8
1.1.2. Tipología.....	9
1.1.3. Prevalencia.....	10
1.1.4. Perfil del agresor.....	11
1.2. Adecuación del Modelo general de clase latente al problema antes descrito.....	12
1.3. Breve descripción del modelo de clase latente.....	16
2. Objetivos del trabajo.....	17
3. El Análisis de Clase Latente.....	18
4. El Análisis de Clase Latente aplicado a medir la calidad de un test binario.....	20
5. Descripción de la Macro-Encuesta de Violencia de Género del CIS de 2019...	22
6. Software de R para llevar a cabo el ajuste del modelo de clase latente.....	27
7. Estimación de la sensibilidad, especificidad y prevalencia de cada uno de los ítems de Violencia de Género.....	30
8. Resumen de los resultados y valoración del método de Análisis de Clase Latente.....	33
9. Conclusiones.....	37
10. Bibliografía.....	39
11. Anexo.....	41

ÍNDICE DE FIGURAS

Figura 1. Gráfico circular de los porcentajes de cada violencia.....	31
Figura 2. Gráfico de embudo lo los porcentajes de las violencias.....	31
Figura 3. Gráfico circular de las mujeres que han sufrido violencia	32
Figura 4. Proporción de mujeres en cada clase latente de Violencia de Control.....	49
Figura 5. Proporción de mujeres en cada clase latente de Violencia Económica.....	55
Figura 6. Proporción de mujeres en cada clase latente de Violencia Psicológica.....	63
Figura 7. Proporción de mujeres en cada clase latente de Violencia Física.....	71
Figura 8. Proporción de mujeres en cada clase latente de Violencia Sexual.....	79

ÍNDICE DE TABLAS

Tabla 1. Prevalencias tradicionales.....	30
Tabla 2. Prevalencias LCA.....	34

1. Introducción

Este trabajo trata sobre la compleja temática de la violencia contra las mujeres (violencia de género), y sus métodos de medición. La violencia de género constituye una grave problemática social que ha persistido a lo largo de los siglos y continúa siendo una preocupación relevante en la actualidad, reflejando la más extrema manifestación de desigualdad y opresión social. Este fenómeno se clasifica en diversas formas, incluyendo la violencia física, psicológica, emocional, económica y sexual, cada una con sus propias características y consecuencias (ONU Mujeres, 2023). La prevalencia de esta violencia se mide mediante estudios epidemiológicos que proporcionan datos esenciales sobre la proporción de mujeres afectadas en un momento dado, destacando la magnitud de este problema social. Adicionalmente, se examina el perfil del agresor, identificando características comunes que permiten una mejor comprensión y abordaje de este problema. Definir claramente el concepto de violencia de género y sus características específicas, como la tipología, la prevalencia y el perfil del agresor, es esencial para llevar a cabo cualquier investigación o intervención con una comprensión más precisa y sin ambigüedades.

En el primer capítulo hablaré de la definición de violencia de género y de cada una de las violencias, tanto individual como específicamente. También trataré de explicar lo que es la prevalencia, cómo se calcula para que nos sirva de ayuda y las diferentes prevalencias que existen. Otro punto importante que abordaré son los distintos perfiles de los agresores, tanto sus características más relevantes como sus perfiles psicológicos.

Por otro lado, introduciré el concepto de análisis de clases latentes (LCA) como una herramienta poderosa para desvelar patrones ocultos y subgrupos no observados dentro de una población. Explicaré los fundamentos del LCA, sus ventajas y aplicaciones en el ámbito de la investigación social, particularmente en el estudio de la violencia de género. El LCA es una técnica estadística avanzada que permite descubrir grupos no observables (clases latentes) dentro de una población a partir de datos observados. A diferencia de métodos tradicionales, el LCA no requiere conocer de antemano a qué grupo pertenece cada individuo. Aplicaré el LCA al estudio de la violencia de género de España en 2019. Utilizaré datos de la macroencuesta nacional sobre violencia de género para identificar clases latentes de mujeres en función de sus experiencias de violencia.

Para finalizar este primer apartado, estudiaré la determinación de los parámetros de calidad diagnóstica de los ítems de las escalas de violencia de género desde la perspectiva de la clase latente.

En el capítulo dos trataré los objetivos de mi trabajo. El objetivo principal es estimar con precisión la prevalencia de la violencia de género e identificar subgrupos de mujeres con diferentes patrones de victimización, utilizando datos de la Macro-Encuesta nacional realizada en 2019. De violencia de género

En el capítulo tres trataré el Modelo de Clases Latentes (LCA), una técnica estadística utilizada para identificar subgrupos no observables dentro de una población a partir de patrones en datos categóricos.

En el capítulo cuatro hablaré sobre el LCA aplicado a medir la calidad de un test binario, junto con un ejemplo. También hablaré de la sensibilidad, la especificidad y del procedimiento a seguir en el LCA.

En el capítulo cinco describiré la Macro-Encuesta de violencia de género de 2019 realizada en España por el CIS.

Siguiendo, en el capítulo seis explico el Software que he utilizado para llevar a cabo el análisis.

El capítulo siete habla sobre la sensibilidad, especificidad y prevalencia de cada uno de los *ítems* y de la violencia en general.

A continuación, en el capítulo ocho explico los resultados obtenidos con el LCA y los discuto con los del análisis estándar.

Finalmente, en el capítulo nueve concluyo con todo lo aprendido y visto en este trabajo.

1.1. Qué es y cómo se mide la violencia contra las mujeres (violencia de género)

1.1.1. Definición

La violencia de género, en términos generales, se describe como cualquier forma de violencia ejercida contra las mujeres simplemente por su condición de género. Esto incluye cualquier acto que busque mantener control sobre las mujeres o imponer una condición de subordinación. Esta violencia es la manifestación más extrema de la desigualdad y opresión de género. Además, es importante

subrayar que esta forma de violencia tiene un carácter social, no siendo atribuible a factores biológicos o psicológicos masculinos, sino a mecanismos sociales que sostienen la subordinación de las mujeres basándose en la diferencia sexual.

1.1.2. Tipología

Según la ONU Mujeres (2023), la violencia contra las mujeres en el ámbito privado abarca diversas formas, incluyendo la violencia económica, psicológica, emocional, física y sexual. A continuación, se describen cada una de estas formas de violencia:

- **Violencia física:** incluye cualquier acción que cause o intente causar daño físico a la mujer.
- **Violencia psicológica:** se refiere a acciones destinadas a degradar o controlar a la mujer.
- **Violencia emocional:** busca específicamente minar la autoestima de la mujer.
- **Violencia económica:** implica controlar los recursos económicos de la mujer, negarles el acceso a ellos o prohibirles trabajar o estudiar.
- **Violencia sexual:** se define como cualquier acto de naturaleza sexual realizado sin el consentimiento de la mujer o cuando esta no puede otorgarlo debido a su edad, discapacidad mental, o estado de intoxicación o inconsciencia. Dentro de la violencia sexual se distinguen:
 - Acoso sexual: incluye contactos físicos no consentidos. También abarca violencia no física, como comentarios sexuales, solicitudes de favores sexuales, miradas insinuantes, acecho, o exhibición de órganos sexuales.
 - Violación: implica cualquier penetración vaginal, anal u oral no consentida, utilizando cualquier parte del cuerpo o un objeto.
 - Violación correctiva: es un tipo de violencia sexual dirigida a una persona debido a su orientación sexual o identidad de género.

1.1.3. Prevalencia

La prevalencia es una medida epidemiológica esencial para evaluar la magnitud de un problema de salud en una población específica, calculando la proporción de individuos que presentan un determinado proceso clínico o resultado en un momento dado. En el contexto de la violencia de género, esta medida de

frecuencia ayuda a determinar cuántas mujeres han sido víctimas de este tipo de violencia en un momento o periodo determinado. La prevalencia se determina a través del muestreo representativo de una población definida en un instante específico, incluyendo tanto a los individuos con el problema de interés como a aquellos que no lo tienen, y se expresa como una proporción o porcentaje.

La fórmula general para calcular la prevalencia (P) es, donde:

- “A” es el número de individuos afectados en un momento específico.
- “B” es el número de individuos no afectados en el mismo momento.

$$P = \frac{A}{A + B}$$

Existen dos tipos principales de prevalencia:

- **Prevalencia puntual**

La prevalencia puntual se calcula en el momento del muestreo para cada individuo de la población definida. Por ejemplo, se puede determinar la prevalencia puntual de personas hospitalizadas por un episodio agudo de asma, donde el numerador es el número de pacientes ingresados por asma aguda y el denominador es el total de pacientes hospitalizados en esa institución en el momento del estudio.

$$\text{Prevalencia de un punto} = \frac{\text{Nº de casos presentes de violencia}}{\text{Total de la población estudiada}}$$

- **Prevalencia de periodo**

La prevalencia de periodo se refiere a los casos presentes en cualquier momento durante un periodo específico. Esta se define como la frecuencia de una enfermedad o condición a lo largo de un periodo y se expresa como una proporción que indica la probabilidad de que un individuo sea un caso en cualquier momento dentro de ese periodo. El numerador de la prevalencia de periodo incluye los casos que se desarrollaron antes y durante el periodo de estudio, mientras que el denominador es la población total durante dicho periodo. Generalmente, se utiliza la población correspondiente al punto medio del periodo como denominador para mayor precisión.

$$\text{Prevalencia de periodo} = \frac{N^{\circ} \text{ casos nuevos} + N^{\circ} \text{ casos presentes de violencia}}{\text{Totalidad de la población estudiada}}$$

Es importante destacar que la prevalencia es una medida estática, que no considera el tiempo de seguimiento y se centra en el estado de la condición en un momento específico. La prevalencia puede ser influenciada por la incidencia (I) y la duración promedio de la enfermedad (T). La relación entre prevalencia, incidencia y duración se expresa mediante la fórmula $P = I * T$, lo que implica que la prevalencia aumentará, disminuirá o se mantendrá estable en función de la incidencia y la duración promedio de la condición.

En el contexto de la violencia de género contra la mujer, la prevalencia se refiere a la proporción de mujeres en una población determinada que han experimentado algún tipo de violencia basada en su género. Esta se mide a través de estudios y encuestas que recogen información auto declarada de las mujeres sobre su experiencia de violencia en un periodo específico, expresándose como un porcentaje o proporción de mujeres afectadas en relación con la población total.

1.1.4. Perfil del agresor

El perfil del agresor proporciona una comprensión más profunda de las características comunes entre aquellos que perpetúan la violencia de género, facilitando la creación de estrategias de intervención más efectivas.

Según la Asociación Gallega de Médicos Forenses (Pastor et al., 2009), no existe un único perfil de agresor, ya que los hombres que ejercen violencia de género conforman un grupo heterogéneo con diversos perfiles psicopatológicos. No obstante, se han identificado ciertas características psicológicas comunes entre los agresores. También Echeburúa (2006) señala que los agresores suelen ser hombres celosos, posesivos, con alta irritabilidad ante los límites, poco control sobre sus impulsos, antecedentes de maltrato y baja autoestima. Por otro lado, el "Estudio medicolegal de los agresores en la violencia de género" realizado por la Facultad de Medicina de la Universidad de Valladolid también encontró características comunes entre los maltratadores. Estas incluyen: Conductas de control y abuso de sustancias, pensamientos extremistas y externalización de la

responsabilidad. Por último, Baja autoestima y escasa asertividad (Tejedor Toquero, 2019).

1.2. Adecuación del Modelo general de clase latente al problema antes descrito

Introducción al Modelo de Clase Latente:

El análisis de clase latente (LCA) es un método estadístico que busca identificar subgrupos no observables dentro de una población basándose en patrones de respuestas en datos categóricos. Este modelo se fundamenta en la premisa de que existe una estructura subyacente en los datos que puede explicar las asociaciones entre las variables observadas, como respuestas a ítems de encuestas o indicadores de comportamiento. Las clases latentes son grupos de individuos que comparten características similares, pero no son directamente observables, y se identifican asignando probabilidades a cada sujeto para pertenecer a cada clase latente según sus patrones de respuestas.

Utilizando un enfoque probabilístico, el LCA estima la probabilidad de pertenencia de cada individuo a las clases latentes, las cuales tienen patrones específicos de respuestas en las variables observadas. Los parámetros del modelo, como las probabilidades de pertenencia a cada clase y las probabilidades de respuestas dentro de cada clase, se estiman mediante métodos estadísticos como el algoritmo de Expectación-Maximización (EM). Una vez ajustado el modelo, las clases latentes identificadas permiten una interpretación más profunda de las estructuras subyacentes en los datos.

En el contexto de la violencia de género, el LCA podría revelar distintos perfiles de víctimas y tipos de violencia que no son evidentes a partir de las respuestas individuales. Esto podría incluir la identificación de grupos de mujeres que experimentan predominantemente violencia económica y psicológica frente a grupos que sufren violencia física y sexual, por ejemplo. Estos hallazgos proporcionan una comprensión más detallada de las distintas experiencias de violencia y permiten desarrollar intervenciones más específicas y efectivas.

Relevancia del Modelo para el Estudio de la Violencia de Género:

El análisis de clase latente (LCA) emerge como una herramienta poderosa y pertinente para abordar el estudio de la violencia de género debido a su capacidad

única para identificar patrones ocultos en datos complejos. En el contexto de la violencia de género, donde las experiencias de las víctimas pueden variar ampliamente y no siempre se reflejan claramente en las respuestas individuales, el LCA permite discernir subgrupos de víctimas y tipos de violencia que podrían estar entrelazados de manera no obvia. Esto es crucial porque la violencia de género abarca una variedad de formas (física, psicológica, económica, sexual, entre otras) y afecta a las víctimas de manera diferenciada, lo que no siempre es captado por enfoques estadísticos convencionales.

Al asignar probabilidades a cada individuo para pertenecer a diferentes clases latentes basadas en patrones de respuestas, el LCA revela estructuras subyacentes que pueden ser fundamentales para comprender las dinámicas complejas de la violencia de género. Por ejemplo, podría identificar grupos de mujeres que experimentan predominantemente violencia económica y psicológica en lugar de formas más visibles como la violencia física. Estos hallazgos no solo ofrecen una visión más completa de las diversas formas de violencia, sino que también permiten diseñar intervenciones más específicas y efectivas. En última instancia, el LCA capacita a los investigadores y profesionales para abordar la violencia de género de manera más holística y centrada en las necesidades reales de las víctimas, proporcionando un marco analítico robusto que puede ampliar significativamente la comprensión y respuesta a este grave problema social.

Aplicación del Modelo al Problema de la Violencia de Género:

El modelo de clase latente (LCA) representa una herramienta fundamental para abordar la complejidad de la violencia de género mediante el análisis de patrones subyacentes en los datos. Este enfoque estadístico permite identificar grupos no observables dentro de una población que comparten características similares en relación con las experiencias de violencia. Por ejemplo, el LCA en violencia de género, puede aplicarse de manera efectiva para descubrir distintos perfiles de víctimas y características de los agresores que no son fácilmente discernibles mediante métodos de análisis convencionales. Para usar el modelo de clase latente a datos de violencia de género, se seleccionan variables observadas que representen diversas dimensiones de este fenómeno. Estas variables pueden incluir diferentes tipos de violencia experimentada por la víctima (física, psicológica, económica, sexual), características del agresor (edad, género, relación con la víctima), y características de la víctima (edad, nivel educativo, situación socioeconómica). Cada una de estas variables aporta información valiosa sobre las interacciones y dinámicas

involucradas en los malos tratos a las mujeres, permitiendo así una exploración más profunda de los diferentes patrones de violencia y sus implicaciones.

Así, mediante el LCA se podría identificar un grupo de mujeres que predominantemente experimentan violencia física y psicológica por parte de un agresor íntimo, mientras que otro grupo podría estar compuesto por mujeres que sufren principalmente violencia económica y sexual en contextos no íntimos. Estos subgrupos proporcionan una comprensión más matizada de las experiencias de las víctimas, ayudando a diseñar intervenciones y políticas más dirigidas y efectivas.

Ventajas del Modelo en este Contexto:

El modelo de clase latente (LCA) ofrece ventajas significativas en la investigación sobre la violencia de género, proporcionando herramientas analíticas que mejoran la comprensión y abordaje de este fenómeno complejo. Una de las principales ventajas del LCA es su capacidad para identificar patrones subyacentes de violencia que pueden no ser evidentes con métodos tradicionales de análisis. Esto permite no solo clasificar a las víctimas en grupos homogéneos según sus experiencias de violencia, sino también entender las dinámicas específicas que caracterizan cada grupo.

Además, el LCA mejora la precisión de las estimaciones de prevalencia al considerar simultáneamente múltiples dimensiones de la violencia de género. Al modelar las interacciones entre diferentes tipos de violencia (física, psicológica, económica, sexual) y las características de los agresores y víctimas, el LCA ofrece insights más detallados sobre cómo estas variables se relacionan entre sí. Por ejemplo, puede revelar cómo ciertos perfiles de agresores están asociados con patrones específicos de violencia hacia las víctimas, contribuyendo así a una comprensión más holística de las dinámicas de poder y control implicadas en los malos tratos a las mujeres.

Limitaciones y Consideraciones:

El modelo de clase latente (LCA) presenta ciertas limitaciones que deben considerarse al aplicarlo al estudio de la violencia de género, junto con aspectos críticos en la interpretación de sus resultados:

Una de las limitaciones principales del LCA es la sensibilidad al número de clases especificadas inicialmente. La determinación incorrecta del número de clases puede llevar a soluciones malas o malinterpretaciones de los patrones de violencia identificados. Esta limitación puede mitigarse mediante técnicas como análisis de

sensibilidad y validación cruzada, que ayudan a evaluar la estabilidad de las soluciones clasificatorias y la robustez de los resultados.

Otra consideración clave es la calidad y completitud de los datos disponibles. Los análisis de clase latente dependen de datos precisos y exhaustivos sobre diferentes formas de violencia de género, así como características detalladas de las víctimas y agresores. La falta de información o datos incompletos pueden sesgar las conclusiones del modelo y limitar su capacidad para capturar la complejidad total de las dinámicas de violencia.

Al interpretar los resultados del LCA en estudios de violencia de género, es esencial considerar la interpretación probabilística inherente al modelo. Las clases latentes identificadas representan perfiles estadísticos y no necesariamente corresponden a grupos homogéneos en la realidad. Por lo tanto, es crucial realizar una validación cualitativa de las clases identificadas y contextualizar los hallazgos dentro del marco teórico y empírico existente sobre violencia de género.

Además, la interpretación de las asociaciones entre variables dentro de cada clase debe realizarse con cautela, reconociendo las complejidades interpersonales y contextuales que influyen en las relaciones abusivas. La combinación del análisis cuantitativo con enfoques cualitativos puede enriquecer la comprensión de los resultados y proporcionar *insights* más profundos sobre las causas y consecuencias de la violencia de género.

Por lo que, mientras el LCA ofrece herramientas poderosas para el estudio de la violencia de género, su aplicación efectiva requiere abordar cuidadosamente las limitaciones metodológicas y considerar de manera crítica la interpretación de los resultados para informar adecuadamente.

1.3. Breve descripción del modelo de clase latente

El análisis de clase latente (LCA) es una técnica estadística avanzada diseñada para descubrir patrones subyacentes o clases latentes dentro de una población a partir de datos observados. A diferencia de los métodos tradicionales, el LCA parte de la premisa de que las respuestas observadas en los individuos son manifestaciones de la pertenencia a clases no directamente observables. Estas clases latentes explican las relaciones entre las variables observadas y permiten clasificar a los individuos según perfiles de respuesta específicos, basándose en principios estadísticos y probabilísticos que subyacen en la estructura de los datos.

El LCA tiene aplicaciones amplias en disciplinas como las ciencias sociales, psicología, medicina, educación y salud pública, por ejemplo. Esta metodología es especialmente valorada por su capacidad para revelar estructuras ocultas en datos complejos, proporcionando *insights* que no son evidentes mediante otros métodos analíticos. Al identificar patrones no obvios y diferencias en las experiencias individuales, el LCA facilita la comprensión profunda de fenómenos complejos y la formulación de estrategias adecuadas para abordarlos.

Pasos del modelado de LCA:

- *Selección de Variables Observadas:* Se eligen variables que capturan aspectos relevantes del fenómeno estudiado, como respuestas a encuestas, características demográficas o comportamientos.
- *Especificación del Modelo:* Se determina el número de clases latentes que se suponen existen en la población, a menudo mediante métodos que evalúan la robustez de diferentes configuraciones de clases.
- *Estimación de Parámetros:* Mediante técnicas como el algoritmo de Expectación-Maximización (EM), se estiman los parámetros del modelo, incluyendo las probabilidades de pertenencia a cada clase y las probabilidades de respuestas en las variables observadas dentro de cada clase.
- *Interpretación de Resultados:* Una vez ajustado el modelo, se interpretan las clases latentes identificadas como perfiles estadísticos que representan grupos de individuos con características similares en relación con las variables observadas.

2. Objetivos del trabajo

En el presente trabajo se propone el análisis de modelos de clase latente (LCA) para abordar la problemática de los malos tratos a mujeres. El objetivo principal es estimar de manera precisa la prevalencia de esta problemática, superando las limitaciones de métodos tradicionales, e identificar subgrupos de mujeres con diferentes patrones de victimización. Para ello, uso los datos de la encuesta nacional sobre violencia contra la mujer realizada en 2019 por el Ministerio de Igualdad y de Asuntos Sociales.

Los LCA permitirán caracterizar subgrupos de mujeres con características similares en las variables observadas (tipo de maltrato, frecuencia, intensidad), comprender mejor la complejidad de la violencia contra la mujer y evaluar la capacidad diagnóstica de los ítems de la escala de violencia utilizada en la encuesta.

Una vez ajustado el modelo de clases latentes a los datos disponibles, se procederá a un análisis detallado de los resultados. Este análisis valorará la estimación de la prevalencia y las estimaciones de la sensibilidad y especificidad de los diferentes indicadores.

Otro objetivo es comparar los resultados obtenidos por LCA, con los del Ministerio que están hechos con un análisis clásico. Se analizarán las características de los subgrupos identificados y se contribuirá a una mejor comprensión de la prevalencia, características y heterogeneidad de los malos tratos a mujeres en el contexto nacional. Se espera que este estudio proporcione información valiosa.

3. El Análisis de Clase Latente

El Análisis de Clases Latentes (LCA) es una técnica estadística fundamental para el estudio de variables categóricas y la identificación de clases latentes. Este método es especialmente útil en las ciencias sociales, donde a menudo se trabaja con datos cualitativos que describen comportamientos, actitudes o características no directamente observables.

Las clases latentes son grupos subyacentes dentro de una población que no se pueden observar directamente, pero que se infieren a partir de patrones en las variables observadas. El LCA asume que las asociaciones entre las variables observadas pueden explicarse mediante una o más variables latentes categóricas, las cuales agrupan a los individuos en clases homogéneas. Esto ayuda a comprender mejor las estructuras subyacentes en los datos y facilita la identificación de subgrupos dentro de una población.

Una comparación común del LCA es con el análisis factorial, ya que ambos buscan identificar estructuras latentes en los datos. Sin embargo, mientras que el análisis factorial se aplica principalmente a datos continuos y busca identificar factores latentes, el LCA se enfoca en datos categóricos y busca identificar clases latentes. Esta distinción hace que el LCA sea particularmente útil para estudios donde las variables observadas son dicotómicas o multinomiales.

El LCA tiene diversas aplicaciones en la investigación empírica. Por ejemplo, se ha utilizado para clasificar tipos de comportamiento del consumidor, segmentar mercados, identificar subgrupos de pacientes con diferentes perfiles de salud mental y explorar patrones de creencias y actitudes en estudios sociológicos. Estas aplicaciones demuestran la versatilidad del LCA para manejar diferentes tipos de datos y contextos de investigación.

El proceso de LCA incluye varios pasos clave: la formulación del modelo, la selección del número de clases latentes, la estimación de parámetros y la evaluación del ajuste del modelo. La formulación del modelo implica definir las variables observadas y las relaciones entre ellas. La selección del número de clases latentes es un paso crítico que a menudo se basa en criterios estadísticos como el AIC (Criterio de Información de Akaike) y el BIC (Criterio de Información Bayesiano). La estimación de parámetros generalmente se realiza mediante métodos de máxima verosimilitud, y la evaluación del ajuste del modelo se lleva a cabo mediante pruebas de bondad de ajuste y análisis de residuos.

En el análisis de grupo único, se asume que todos los individuos provienen de una población homogénea y se busca identificar las clases latentes dentro de esa población. Este enfoque es útil cuando se trabaja con una muestra relativamente uniforme o cuando el objetivo es explorar la estructura latente dentro de un grupo específico.

En el análisis de grupo múltiple, se considera que los individuos pueden provenir de diferentes subpoblaciones. Esto permite comparar las clases latentes entre diferentes grupos y explorar cómo las estructuras latentes pueden variar según factores como la edad, el género, la ubicación geográfica u otras variables relevantes. En *Haagenars & McCutcheon* (2002) se explican cómo extender los modelos de LCA para manejar múltiples grupos y cómo interpretar las diferencias y similitudes entre las clases latentes identificadas en cada grupo.

La estimación de parámetros en LCA destaca el uso de algoritmos de máxima verosimilitud y EM (*Expectation-Maximization*). Estos métodos permiten estimar las probabilidades de pertenencia a cada clase latente y las probabilidades condicionales de las respuestas observadas dentro de cada clase. Además, se discuten diversas técnicas para evaluar el ajuste del modelo, incluyendo la revisión de los índices de ajuste y la comparación de modelos alternativos.

4. El Análisis de Clase Latente aplicado a medir la calidad de un test binario

Presentación de los parámetros de un test diagnóstico

En el ámbito de los test diagnósticos, es fundamental definir y comprender los parámetros que miden la calidad del test. Los principales parámetros son la sensibilidad y la especificidad. Según el experto en estadística y metodología de la investigación, Douglas G. Altman, la sensibilidad se refiere a “la capacidad de una prueba para identificar correctamente a los individuos que tienen una enfermedad”, ya sea enfermedad o condición (verdaderos positivos), mientras que la especificidad se refiere a “la capacidad de una prueba para identificar correctamente a los individuos que no tienen la enfermedad”, enfermedad o condición (verdaderos negativos) (Altman, 1991).

El "*gold standard*" es la prueba más confiable para medir algo, como una enfermedad. Sirve para comparar otras pruebas y ver qué tan precisas son. Es importante porque nos ayuda a saber si una prueba nueva es buena o no. Sin embargo, no siempre existe una prueba perfecta como *gold standard*, y algunas pueden tener errores.

Cuando se dispone de un estándar de referencia o "*gold standard*", estos parámetros pueden estimarse directamente comparando los resultados del test con los resultados del *gold standard* (Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M.,... & de Vet, H. C. (2003)). La sensibilidad se calcula como la proporción de verdaderos positivos sobre el total de individuos que realmente tienen la condición, y la especificidad se calcula como la proporción de verdaderos negativos sobre el total de individuos que no tienen la condición:

$$\text{Sensibilidad} = \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{Falsos negativos}}$$

y

$$\text{Especificidad} = \frac{\text{Verdaderos Negativos}}{\text{Verdaderos Negativos} + \text{Falsos Positivos}}$$

Estimación de parámetros en ausencia de *gold_standard*

En muchos casos, no se dispone de un *gold_standard* infalible. En estas situaciones, el *Análisis de Clase Latente* (LCA) se convierte en una herramienta poderosa para estimar la calidad de los test diagnósticos. El LCA permite modelar la relación entre múltiples

test diagnósticos y una variable latente no observada que representa la verdadera condición del individuo.

El modelo LCA asume que las asociaciones entre los test observados pueden explicarse mediante una o más variables latentes categóricas, agrupando a los individuos en clases homogéneas. Esto facilita la estimación de la sensibilidad y especificidad de los test sin necesidad de un *gold_standard*. Los parámetros del modelo se estiman generalmente mediante el método de máxima verosimilitud o utilizando enfoques bayesianos.

Aplicación del LCA en la evaluación de test diagnósticos

Para ilustrar la aplicación del LCA en la evaluación de test diagnósticos, considero un estudio “realizado en Tanzania para detectar leptospirosis en pacientes febriles”(Matthew R. Schofield, Michael J. Maze², John A. Crump, Matthew P. Rubach, Renee Galloway, Katrina J. Sharples(2021)), descrito en el artículo sobre robustez de modelos LCA. En este estudio, se utilizaron varios test diagnósticos, incluidos un test estándar y tres nuevos test de punto de atención. Dado que no se disponía de un *gold_standard* infalible, se empleó un modelo de clases latentes bayesiano para evaluar las propiedades de estos nuevos test.

El modelo LCA permitió estimar las sensibilidades y especificidades de los test, revelando que los nuevos test mostraban sensibilidades más altas que el test estándar, aunque con mayor incertidumbre. Este enfoque también permitió evaluar la consistencia de las estimaciones y la adecuación del modelo mediante la evaluación predictiva posterior.

Procedimiento del LCA

Pasos importantes del proceso del LCA:

- *Formulación del Modelo*: Definir las variables observadas y las relaciones entre ellas.
- *Selección del Número de Clases Latentes*: Determinar el número de clases latentes utilizando criterios estadísticos como el AIC (Criterio de Información de Akaike) y el BIC (Criterio de Información Bayesiano).
- *Estimación de Parámetros*: Generalmente se realiza mediante métodos de máxima verosimilitud o algoritmos de Expectación-Maximización (EM).
- *Evaluación del Ajuste del Modelo*: Realizar pruebas de bondad de ajuste y análisis de residuos para verificar la adecuación del modelo.

5. Descripción de la Macro-Encuesta de Violencia de Género del CIS de 2019

La *Macroencuesta de Violencia contra la Mujer 2019* fue realizada por el Centro de Investigaciones Sociológicas (CIS) en colaboración con la Delegación del Gobierno contra la Violencia de Género (Delegación del Gobierno contra la Violencia de Género & CIS, 2019). Este estudio, que se enmarca en la serie de macroencuestas sobre violencia contra la mujer que el CIS lleva a cabo desde 2007, se realiza cada cuatro años y tiene como objetivo principal estimar la prevalencia de la violencia contra la mujer en España, tanto en el ámbito de la pareja o expareja como fuera del mismo.

La encuesta se realizó a una muestra representativa de 9.568 mujeres mayores de 16 años residentes en España. La recogida de datos se llevó a cabo entre los meses de marzo y mayo de 2019 mediante entrevistas telefónicas asistidas por ordenador.

El cuestionario consta de 1.121 preguntas. Consta de cuatro grandes bloques:

- El primero se enfoca en el análisis de la violencia en la pareja actual(M1) o exparejas(M2). Midiendo las violencias: física, sexual, psicológica emocional, psicológica de control, económica, miedo de la pareja, consecuencias y denuncias.
- El segundo se enfocará en el análisis de la violencia física y sexual fuera de la pareja(M3).
- El tercero se basa en el estudio del acoso sexual y del acoso reiterado.
- El cuarto se centra en el análisis de lo que sucede en mujeres especialmente vulnerables: mujeres con discapacidad, jóvenes de 16 a 24 años, mujeres de 65 o más años, mujeres que han nacido en el extranjero, y mujeres que viven en municipios pequeños.
- Finalmente, con este análisis se pretendía proporcionar una visión detallada y actual de la extensión y las formas de la violencia contra la mujer, así como de distintos aspectos relacionados con la misma.

Sus principales objetivos son conocer el porcentaje de mujeres de 16 o más años residentes en España que han sufrido o sufren algún tipo de violencia por ser mujeres. También conocer los tipos de agresores más frecuentes.

Por otro lado, voy a mencionar para cada violencia en qué items se basan:

Violencia física en pareja (M1P4, M1P4a, M1P4b, M2P4, M2P4a, M2P4b, M3P1):

En total, 1016 mujeres han sufrido violencia física.

- Le ha abofeteado o tirado algo que pudiese hacerle daño. 674 mujeres han sufrido esta violencia.
- La ha empujado, agarrado o tirado del pelo. 833 mujeres han sufrido esta violencia.
- La ha golpeado con su puño o con alguna otra cosa que pudiese hacerle daño. 475 mujeres han sufrido esta violencia.
- Le ha dado patadas, arrastrado o pegado una paliza. 369 mujeres han sufrido esta violencia.
- La ha intentado asfixiar o quemar a propósito. 205 mujeres han sufrido esta violencia.
- La ha amenazado con usar o ha usado una pistola, cuchillo u otra arma o sustancia peligrosa contra Ud. 213 mujeres han sufrido esta violencia.
- Ha usado la fuerza contra usted, de cualquier otra manera de las mencionadas anteriormente, de forma que le ha hecho daño o podría haberle hecho daño. 406 mujeres han sufrido esta violencia.

Violencia sexual en pareja se preguntaban los ítems: (M1P5, M1P5a, M1P5b, M2P5, M2P5a, M2P5b, M3P2).

En total, 860 mujeres han sufrido violencia sexual.

- La ha obligado a mantener relaciones sexuales amenazándole, sujetándola o haciéndole daño de alguna manera. 405 mujeres han sufrido esta violencia.
- Le ha hecho mantener relaciones sexuales cuando era incapaz de rechazarlas debido a que usted estaba bajo la influencia del alcohol o las drogas. 201 mujeres han sufrido esta violencia.
- Ha mantenido relaciones sexuales sin desearlo porque tenía miedo de lo que le podría hacer si se negaba. 503 mujeres han sufrido esta violencia.
- La ha obligado a mantener relaciones sexuales cuando Ud. no quería. 613 mujeres han sufrido esta violencia.
- Ha intentado obligarla a tener relaciones sexuales contra su voluntad sin conseguirlo. 456 mujeres han sufrido esta violencia.

- Le ha tocado a Ud. sus partes íntimas o le ha realizado algún otro tipo de tocamiento de tipo sexual cuando usted no quería. 526 mujeres han sufrido esta violencia.
- Le ha hecho alguna vez tocarle sus partes íntimas o la ha obligado a realizarle algún otro tipo de tocamiento de tipo sexual cuando usted no quería. 392 mujeres han sufrido esta violencia.
- La ha obligado a realizar alguna otra práctica de tipo sexual que yo no le haya mencionado ya. 242 mujeres han sufrido esta violencia.

Violencia psicológica emocional en la pareja Se preguntan los siguientes ítems: (M1P3, M1P3a, M1P3b, M2P3, M2P3a, M2P3b):

En total, 2185 mujeres han sufrido violencia psicológica emocional.

- La ha insultado o hecho sentirse mal con usted misma. 1822 mujeres han sufrido esta violencia.
- La ha menospreciado o humillado delante de otras personas. 1352 mujeres han sufrido esta violencia.
- La ha asustado o intimidado a propósito. 1223 mujeres han sufrido esta violencia.
- La ha amenazado verbalmente con hacerle daño a Ud. 878 mujeres han sufrido esta violencia.
- La ha amenazado verbalmente con hacerle daño a sus hijos/as o a alguna otra persona que es/era importante para Ud. 427 mujeres han sufrido esta violencia.
- La ha amenazado con hacerse daño a sí mismo/a si usted lo/la deja. 595 mujeres han sufrido esta violencia.
- La ha amenazado con quitarle a sus hijos/as. 368 mujeres han sufrido esta violencia.

Violencia psicológica de control en la pareja: Preguntan los siguientes ítems: (M1P1, M1P1a, M1P1b, M2P1, M2P1a, M2P1b):

En total, 2516 mujeres han sufrido violencia psicológica de control.

- Trata o ha tratado de impedirle que vea a sus amigos o amigas. 1254 mujeres han sufrido esta violencia.
- Trata o ha tratado de evitar que Ud. se relacione con su familia directa o parientes. 810 mujeres han sufrido esta violencia.

- Insiste o ha insistido en saber dónde está usted en cada momento. 1548 mujeres han sufrido esta violencia.
- La ignora o ha ignorado y la trata o ha tratado con indiferencia. 1433 mujeres han sufrido esta violencia.
- Se enfada o se ha enfadado si habla con otro hombre o mujer. 1502 mujeres han sufrido esta violencia.
- Sospecha o ha sospechado sin motivos que Ud. le es/era infiel. 1175 mujeres han sufrido esta violencia.
- Espera o ha esperado que Ud. le pida permiso antes de ir por su cuenta a determinados sitios. 852 mujeres han sufrido esta violencia.

Violencia económica en la pareja Se preguntan los siguientes ítems: (M1P1, M1P1a, M1P1b, M2P1, M2P1a, M2P1b):

En total, 1113 mujeres han sufrido violencia económica.

- Se niega o se ha negado a darle dinero para los gastos del hogar cuando la pareja tiene/tenía dinero para otras cosas. 693 mujeres han sufrido esta violencia.
- Le impide o ha impedido tomar decisiones relacionadas con la economía familiar y/o realizar las compras de forma independiente. 789 mujeres han sufrido esta violencia.
- No la deja o no la ha dejado trabajar o estudiar fuera del hogar. 160 mujeres han sufrido esta violencia.
- Usa o ha usado su dinero o su tarjeta de crédito o pide préstamos a su nombre sin su consentimiento. 512 mujeres han sufrido esta violencia.

Miedo de la pareja. Se preguntan: (M1P6, M1P6a, M2P6, M2P6a):

- M1P6 ¿Con qué frecuencia ha tenido o tiene Ud. miedo de su pareja actual?
- M2P6 ¿Con qué frecuencia ha tenido o tiene Ud. miedo de alguna de su/s pareja/s del pasado? Siendo las posibles respuestas - Continualmente - Muchas veces - Algunas veces – Nunca

A las que respondían que han sentido miedo continuamente, algunas veces, o muchas veces, se les preguntaba a su vez con qué frecuencia lo habían sentido en los 12 meses previos a las entrevistas siendo de nuevo las posibilidades de respuesta: continuamente, muchas veces, algunas veces, y nunca. En el capítulo 31 se analiza en primer lugar la prevalencia y la frecuencia del miedo en la pareja a lo largo de la vida y en los doce meses

previos a las entrevistas. Posteriormente se estudia hasta qué punto la variable miedo de la pareja permite descubrir casos ocultos de violencia en la pareja.

6. Software de R para llevar a cabo el ajuste del modelo de clase latente

Los datos son sacados de la Delegación del Gobierno contra la Violencia de Género, de una macroencuesta de la violencia contra la mujer realizada en 2019, (Ministerio de Igualdad, 2024).

Al mismo tiempo, los abro en *Rstudio* y los meto en un *dataframe*. Observo los datos y veo qué columnas tienen todas las filas el mismo valor. Elimino las columnas que no me dan información relevante como: FECHAINI, HORAINI, FECHAFIN, HORAFIN, DURACIÓN. También elimino las columnas que todas sus filas tienen el mismo valor como: ESTUDIO, DISTR, SECCIÓN, ENTREV. En total elimino 11 columnas. La versión que utilizo en R es: 2023.06.1+524. Lanzada el 7 de julio de 2023.

A continuación, leo el cuestionario lentamente para ir viendo los “FILTROS” que tiene y representarlos en mis datos de R. Por ejemplo: “Si NO M0P4=1 ir a [M0P5]”. Hago lo mismo con los “SALTOS”, por ejemplo: “Si M0P4A= (2;9) ir a [M0P5]”.

También, en ciertos casos, reemplazo los valores NA (valores ausentes) por 0. Por ejemplo: en el FILTRO “Si NO M0P4A=1 ir a M0P5” en todas las situaciones M0P4A!= 1 hasta M0P5, entonces paso todos los valores de la fila NA a 0.

Las librerías que he utilizado son:

1. *Haven*:

Es una librería en R diseñada para la importación y exportación de datos desde y hacia otros formatos de software estadístico, especialmente archivos de datos de SPSS, Stata y SAS. Esta librería proporciona funciones para leer datos de estos formatos directamente en R, manteniendo los metadatos (como etiquetas de variables y valores perdidos) que pueden estar presentes en estos archivos. He usado la función `read_spss` para leer un archivo de SPSS en R.

2. *Lavaan*:

Es una librería utilizada para la especificación y estimación de modelos de ecuaciones estructurales (SEM), incluyendo el análisis de clase latente (LCA). SEM es una técnica estadística poderosa que permite modelar relaciones complejas entre variables observadas y no observadas. *Lavaan* también es adecuada para el análisis de clases latentes (LCA), que es un tipo específico de SEM utilizado para identificar subgrupos no observados en una población basándose en patrones de respuestas.

3. *Tidyr*:

Es una librería que ofrece herramientas para manipular y transformar datos tabulares de manera eficiente. Tiene funciones para reorganizar datos. Además, facilita la separación y unión de columnas y la expansión de listas. Uso la función *pivot_longer* para convertir los datos de formato ancho a largo y *unite* que combina varias columnas en una sola.

4. *Devtools*:

Es una librería que proporciona herramientas para instalar, crear y gestionar paquetes R, así como para trabajar con repositorios de paquetes (como GitHub).

5. *poLCA*:

Es una librería que implementa funciones para el análisis de modelos de clases latentes con variables ordinales. Este tipo de análisis se utiliza cuando las variables de interés son ordinales y deseas identificar patrones subyacentes o perfiles de respuesta entre los individuos. Uso la función *poLCA* para ajustar modelos de clases latentes.

6. *MASS*:

Es una librería que proporciona un conjunto de funciones estadísticas y herramientas gráficas para el análisis de datos en ciencias sociales y más. Incluye métodos para ajuste de modelos lineales, análisis de componentes principales, *clustering*, entre otros. He usado la función *stepAIC* para seleccionar modelos por el criterio de iAIC.

7. *Readxl*

Es una librería que se utiliza para importar datos desde archivos de Excel (.xls y .xlsx) directamente a R. Es particularmente útil cuando necesitas trabajar con datos almacenados en hojas de cálculo de Excel y deseas cargar esos datos en R para su análisis posterior. Uso la función *read_excel* para leer un documento de Excel en R.

8. *Dplyr*

Es una librería muy poderosa en R que proporciona una gramática consistente para manipular datos. Se centra en las operaciones más comunes de manipulación de datos, como filtrado, selección, reordenamiento, agregación y unión de datos. Esta librería permite realizar estas operaciones de manera eficiente y con una sintaxis intuitiva. Uso las funciones: *select* para seleccionar columnas del *dataframe*. *Mutate* para añadir nuevas variables y transformarlas. Y *across* para aplicar una función a varias columnas.

9. ggplot2

Es una librería de visualización en R que permite crear gráficos de alta calidad de una manera sencilla y declarativa. Utiliza una gramática de gráficos que separa completamente los datos de su representación gráfica, lo que facilita la creación de gráficos complejos con capas y ajustes estéticos. Uso un par de funciones: *ggplot* para crear un nuevo objeto *ggplot*. La función *aes* para establecer las asignaciones estéticas. También *geom_bar* para añadir una capa de gráfico de barras. Por otro lado, la función *labs* para personalizar etiquetas y por último la función *theme_minimal* para aplicar un tema minimalista al gráfico.

7. Estimación de la sensibilidad, especificidad y prevalencia de cada uno de los ítems de Violencia de Género

Para la sensibilidad, especificidad y prevalencia, como es una encuesta, voy a asumir que todas las mujeres han contestado sinceramente; es decir: que los valores 1 indican que han sufrido violencia (Positivo) y 0 indican que no ha sufrido violencia (Negativo). Admito ese error al asumir que los Falsos Negativos son 0 y los Falsos Positivos son 1. Las escalas que voy a tratar son las mencionadas anteriormente, en el capítulo Escala de Control, de Violencia Económica, de Violencia Psicológica de Violencia Física de Violencia Sexual y de Miedo. También trataré la escala conjunta de Violencia Psicológica, Física y Sexual y la escala conjunta de Violencia Física y Sexual. A continuación, vamos a calcular las prevalencias de los datos limpios, sin clases latentes. Para calcular la sensibilidad, especificidad y prevalencia, hay que definir lo siguiente:

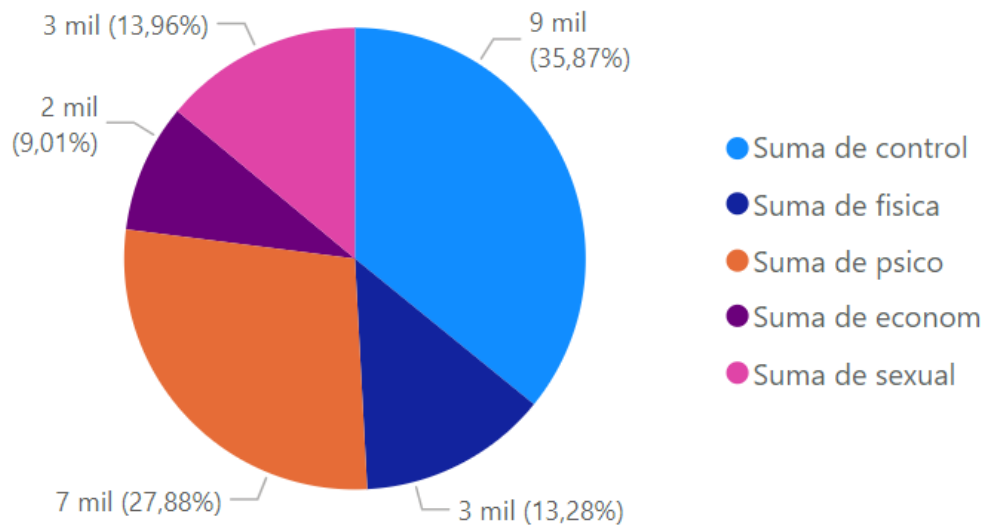
- Verdaderos positivos (TP): han sufrido violencia y es correcto.
- Falsos positivos (FP): han sufrido violencia y es falso.
- Verdaderos negativos (TN): no han sufrido violencia y es correcto.
- Falsos negativos (FN): no han sufrido violencia y es falso.

Tabla 1. Prevalencias Tradicionales

VIOLENCIA	SENSIBILIDAD	ESPECIFICIDAD	PREVALENCIA
Control	1	1	0.26285535
Económica	1	1	0.11622074
Psicológica	1	1	0.22826087
Física	1	1	0.10608278
Sexual	1	1	0.08977843
Física y Sexual	1	1	0.14025920
Psicológica, Física y Sexual	1	1	0.24111622
No violencia	1	1	0.6826923
Violencia	1	1	0.31730769

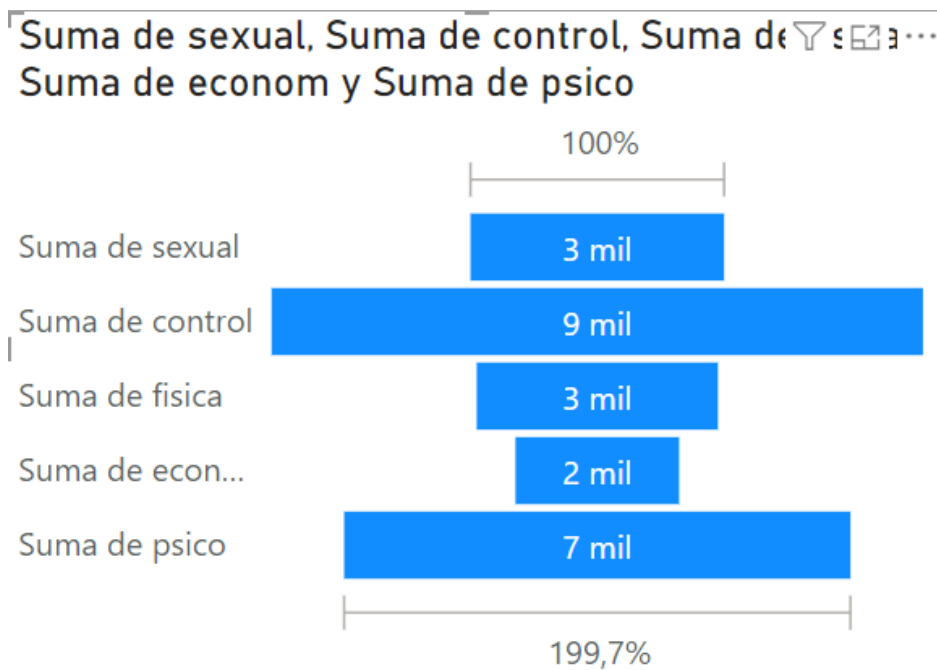
Gráfica 1. Porcentajes cada violencia

Suma de control, Suma de fisica, Suma de psico, Suma de econom y Suma de sexual



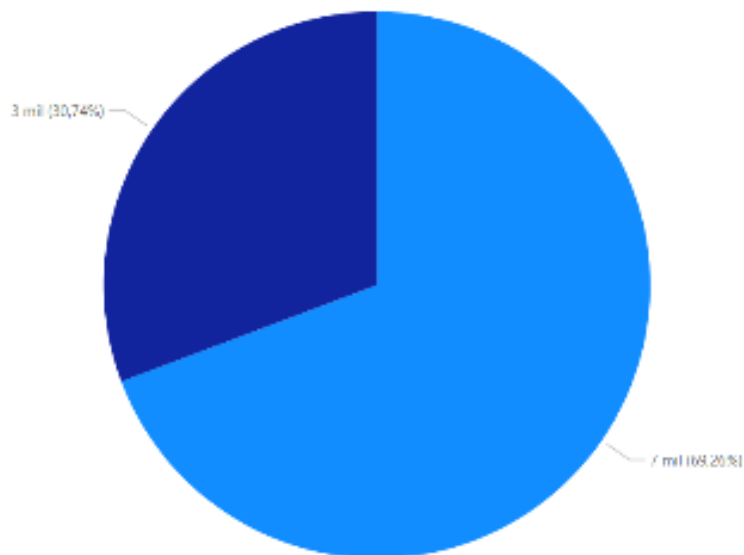
En la gráfica 1, podemos observar cuántas mujeres han sufrido cada una de las violencias; siendo la de control la que más veces se ha sufrido y la violencia sexual la que menos. Esto mismo reflejaba la tabla 1.

Gráfica 2. Gráfica embudo de los porcentajes de cada violencia.



En esta segunda gráfica en forma de embudo, se puede ver mejor representada la cantidad de distintos tipos de violencias y sus diferencias.

Gráfica 3. Porcentaje de mujeres que han sufrido violencia



En la gráfica 3 vemos de azul oscuro las mujeres que han sufrido algún tipo de violencia (31.73%) y de azul claro las que no han sufrido ninguna de estas (68.27%).

Viendo estas figuras podemos concluir que hay casi un tercio de las mujeres que sufren algún tipo de violencia. Dentro de estas, la más sufrida es la violencia de control, seguida por la psicológica. Con un porcentaje más leve tenemos la violencia física, sexual y económica.

8. Resumen de los resultados y valoración del método de Análisis de Clase Latente

Este análisis ha revelado varias conclusiones significativas con implicaciones. Para empezar, el número de clases de cada violencia ha sido determinado por las variables AIC y BIC. La variable AIC (*Aikake Information Criterion*) mide la calidad del modelo en función de la pérdida de información, y el modelo BIC (*Bayesian Information Criterion*) es similar al AIC, pero introduce una penalización mayor por el número de parámetros del modelo; cuanto menor es el BIC el ajuste es mejor. A parte de estos valores, también he hecho pruebas de significancia para ver si agregando más o menos clases mejora significativamente el ajuste del modelo. Ahora veremos cada violencia con sus clases:

- **Violencia de Control:** Se identificaron tres clases latentes:
 - **Clase 1:** Incluye al 60% de las mujeres.
 - **Clase 2:** Incluye al 30% de las mujeres.
 - **Clase 3:** Incluye al 10% de las mujeres.
- **Violencia Económica:** Se identificaron dos clases latentes:
 - **Clase 1:** Incluye al 83.28% de las mujeres.
 - **Clase 2:** Incluye al 16.72% de las mujeres.
- **Violencia Psicológica:** Se identificaron tres clases latentes:
 - **Clase 1:** Incluye al 50% de las mujeres.
 - **Clase 2:** Incluye al 30% de las mujeres.
 - **Clase 3:** Incluye al 20% de las mujeres.
- **Violencia Física:** Se identificaron tres clases latentes:
 - **Clase 1:** Incluye al 40% de las mujeres.
 - **Clase 2:** Incluye al 35% de las mujeres.
 - **Clase 3:** Incluye al 25% de las mujeres.
- **Violencia Sexual:** Se identificaron tres clases latentes:
 - **Clase 1:** Incluye al 70% de las mujeres.
 - **Clase 2:** Incluye al 20% de las mujeres.
 - **Clase 3:** Incluye al 10% de las mujeres.

A continuación, vemos las probabilidades condicionales de cada clase en su tipo de violencia, es decir, las probabilidades de cada una de responder "Sí" o "No": Mencionaré los primeros ítems de cada escala.

- **Violencia de Control:**
 - **Clase 1:** 86.89% de probabilidad de responder "Sí".
 - **Clase 2:** 91.53% de probabilidad de responder "Sí".

- **Clase 3:** 83.82% de probabilidad de responder "Sí".
- **Violencia Económica:**
 - **Clase 1:** 98.95% de probabilidad de responder "No".
 - **Clase 2:** 73.18% de probabilidad de responder "Sí".
- **Violencia Psicológica:**
 - **Clase 1:** 99.55% de probabilidad de responder "No".
 - **Clase 2:** 59.47% de probabilidad de responder "No".
 - **Clase 3:** 95.91% de probabilidad de responder "Sí".
- **Violencia Física:**
 - **Clase 1:** 99.55% de probabilidad de responder "No".
 - **Clase 2:** 59.47% de probabilidad de responder "No".
 - **Clase 3:** 95.91% de probabilidad de responder "Sí".
- **Violencia Sexual:**
 - **Clase 1:** 99.55% de probabilidad de responder "No".
 - **Clase 2:** 59.47% de probabilidad de responder "No".
 - **Clase 3:** 95.91% de probabilidad de responder "Sí".

En la siguiente tabla se diferencian las violencias por clases. Se mencionan el porcentaje de mujeres pertenecientes a cada clase y la probabilidad de que respondan sí o no dentro de cada una de ellas:

	Control	Económica	Psicológica	Física	Sexual
% Clase 1	60%	83.28%	50%	40%	70%
% "Sí"	86.89%	1.05%	0.45%	0.45%	0.45%
% "No"	13.11%	98.95%	99.55%	99.55%	99.55%
% Clase 2	30%	16.72%	30%	35%	20%
% "Sí"	91.53%	73.18%	40.53%	40.53%	40.53%
% "No"	8.47%	26.82%	59.47%	59.47%	59.47%
% Clase 3	10%	-	20%	25%	10%
% "Sí"	83.82%	-	95.91%	95.91%	95.91%
% "No"	16.18%	-	4.09%	4.09%	4.09%

Tabla 2. Prevalencias LCA.

Distribución de la Violencia:

- La mayoría de las mujeres se encuentra en subgrupos con baja exposición a la violencia económica. La clase 1 corresponde al 83.28% de las mujeres y éstas tienen una probabilidad de 0.98 de responder "No", y a la violencia sexual, donde la clase 1 son el 70% de las mujeres, tienen la probabilidad de responder "No", misma probabilidad que en la clase 1 de la violencia física y psicológica. Por lo tanto, tenemos que, de todas las escalas estudiadas menos en la violencia de control, en la clases 1, tienen una alta probabilidad de no haber sufrido esta violencia y son un alto porcentaje las que se encuentran en este grupo.
- Sin embargo, una proporción significativa de mujeres experimenta alta exposición a la violencia psicológica y física, perteneciendo a la Clase 3 representando el 20% y el 25%, respectivamente, y con altas probabilidades de responder "Sí" a los ítems correspondientes, 0.9591 en ambas violencias.
- Si nos fijamos en la violencia de Control, como hemos visto en los gráficos del *apartado 7*, es la violencia más ejercida y en todas las Clases tiene una alta probabilidad de haberla sufrido no bajando de un 83.82%.

Las prevalencias estimadas utilizando LCA muestran una distribución más matizada y detallada en comparación con los métodos tradicionales, como podemos ver comparando la *tabla 1* del *apartado 7* permitiendo identificar subgrupos específicos que pueden no ser visibles a través de análisis convencionales.

La fiabilidad del método de LCA se evalúa considerando la robustez y la coherencia de los resultados obtenidos:

- **Robustez del Modelo:**
 - **Consistencia de Resultados:** Las clases latentes identificadas son consistentes y las probabilidades condicionales muestran patrones claros y diferenciados entre las clases, lo que sugiere una robustez en la segmentación de los datos.
 - **Adecuación del Ajuste:** El ajuste del modelo de LCA fue adecuado según los criterios de ajuste (e.j., BIC, AIC), lo que indica que el modelo utilizado es apropiado para los datos analizados.
- **Fiabilidad en la Identificación de Subgrupos:**
 - **Claridad en la Segmentación:** La identificación de subgrupos con diferentes niveles de exposición a diversos tipos de violencia proporciona

una visión detallada y matizada que es crucial para el diseño de intervenciones específicas.

- **Reproducibilidad:** Los resultados obtenidos pueden ser reproducibles en estudios similares, lo que incrementa la fiabilidad de los hallazgos.

- **Comparación con Estudios Previos:**

- **Consistencia con la Literatura:** Los hallazgos del análisis de clases latentes son consistentes con estudios previos sobre violencia de género, lo que refuerza la fiabilidad del método.

Los resultados sobre la robustez del modelo, la fiabilidad en la identificación de subgrupos y la consistencia con estudios previos sobre violencia de género se alinean con lo reportado en el estudio publicado en “*Statistics in Medicine*” (2021), donde se evaluaron criterios de ajuste como BIC y AIC para validar la adecuación del modelo LCA así como la claridad y reproducibilidad de la segmentación de datos de la violencia de género.

9. Conclusiones

El análisis de clases latentes ha demostrado ser un método fiable y robusto para analizar la violencia de género en España. Ha permitido identificar subgrupos específicos y detallados, proporcionando información crucial para la formulación de políticas y programas de intervención. La fiabilidad de los resultados, apoyada por la consistencia y robustez del modelo, sugiere que LCA es una herramienta valiosa para estudios futuros y la evaluación continua de la violencia de género.

Este estudio ha proporcionado una visión integral de la violencia de género en base a los Modelos de Clase Latente (LCA), lo que ha permitido superar las limitaciones de los métodos tradicionales y ofrecer una estimación más precisa de la prevalencia de la violencia contra las mujeres. A continuación, se presentan las principales conclusiones del análisis.

Los Modelos de Clase Latente han permitido identificar subgrupos de mujeres con diferentes patrones de victimización. Por ejemplo, he encontrado tres clases latentes para la violencia de control, económica, psicológica, física y sexual, cada una con características y prevalencias distintas.

La mayoría de las mujeres se agrupan en clases con baja exposición a la violencia económica y sexual, mientras que una proporción significativa experimenta alta exposición a la violencia psicológica y física.

La aplicación del LCA mejora en la Estimación de la Prevalencia. Ha mostrado una distribución más matizada y detallada de las prevalencias en comparación con los métodos tradicionales. Esto ha permitido identificar subgrupos específicos que pueden no ser visibles a través de análisis convencionales. Las clases latentes identificadas muestran patrones claros y diferenciados entre las clases, lo que sugiere una segmentación robusta y precisa de los datos.

Los resultados obtenidos son consistentes y reproducibles, lo que aumenta la fiabilidad de los hallazgos. La adecuación del ajuste del modelo de LCA, evaluado mediante criterios como BIC y AIC, ha sido confirmada. La fiabilidad en la identificación de subgrupos ha sido destacada, proporcionando una visión detallada y matizada crucial para el diseño de intervenciones específicas.

La comparación entre los resultados obtenidos por LCA y los análisis clásicos del Ministerio ha mostrado que los Modelos de Clase Latente ofrecen una comprensión más profunda de la prevalencia, características y heterogeneidad de los malos tratos a

mujeres. El LCA ha demostrado ser una herramienta valiosa y eficaz para identificar subgrupos de mujeres que no son evidentes en los análisis convencionales, lo que refuerza la relevancia y ventajas de este método. La identificación de subgrupos específicos con diferentes niveles de exposición a la violencia puede guiar el diseño de intervenciones más efectivas y personalizadas.

En resumen, este trabajo ha demostrado que el Análisis de Clase Latente es un método potente y eficaz para abordar la problemática de la violencia de género, ofreciendo estimaciones precisas y detalladas que pueden mejorar significativamente la comprensión y abordaje de esta grave problemática social.

En un futuro, se podría ampliar el análisis a otras encuestas y contextos para verificar la reproducibilidad y consistencia de los resultados.

Bibliografía

1. Altman, D. G. (1991). *Practical Statistics for Medical Research*. Chapman & Hall.
2. Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., ... & de Vet, H. C. (2003). The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clinical Chemistry*, 49(1), 7-18.
3. Castro, R. (2016). Violencia de género. *Conceptos clave en los estudios de género*, 1, 339-354.
4. Collins, L. M., & Lanza, S. T. (2010). *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. John Wiley & Sons.
5. Delegación del Gobierno contra la Violencia de Género, & Centro de Investigaciones Sociológicas. (2019). *Macroencuesta de Violencia contra la Mujer 2019*. Ministerio de Igualdad.
6. Echeburúa, E. (2006). Hombres violentos contra la pareja: perfil psicopatológico y programa de intervención. *Formación continuada para Médicos Forenses*. Centro de Estudios Jurídicos de la Administración de Justicia. Ministerio de Justicia.
7. Fajardo-Gutiérrez, A. (2017). Medición en epidemiología: prevalencia, incidencia, riesgo, medidas de impacto. *Revista Alegría México*, 64(1), 109-120. DOI: 10.29262/ram.v64i1.252
8. Ferrer, M. E. F., & Del Prado González, N. (2013). Medidas de frecuencia y de asociación en epidemiología clínica. *Canales de Pediatría Continuada*, 11(6), 346-349. DOI: 10.1016/s1696-2818(13)70157-4
9. Hagenaars, J. A., & McCutcheon, A. L. (Eds.). (2002). *Applied Latent Class Analysis*. Cambridge University Press.
10. Lanza, S. T., Collins, L. M., Lemmon, D. R., & Schafer, J. L. (2007). PROC LCA: A SAS procedure for latent class analysis. *Structural Equation Modeling*, 14(4), 671-694.
11. Latent class Analysis | *MPlus Data Analysis Examples*. (n.d.). Retrieved from <https://stats.oarc.ucla.edu/mplus/dae/latent-class-analysis/>
12. Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent Structure Analysis*. Houghton Mifflin.
13. Linzer, D. A., & Lewis, J. (2013). poLCA: *Polytomous variable latent class analysis* (R package version 1.4). Retrieved from <https://dlinzer.github.com/poLCA>
14. Magidson, J., & Vermunt, J. K. (2004). Latent class models. En D. Kaplan (Ed.), *The Sage Handbook of Quantitative Methodology for the Social Sciences* (pp. 175-198). Sage Publications.
15. Mplus. (n.d.). Documentation and examples for performing latent class analysis. Retrieved from Mplus.

16. Muthén, B. O., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, 24(6), 882-891.
17. NIUS. (n.d.). ¿Qué tipos de violencia de género existen? NIUS. Recuperado de <https://www.niusdiario.es>
18. Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*, 4(4), 440.
19. Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, 14(4), 535-569.
20. Pastor, Rodes, & Navarro. (2009). Perfil del agresor en la violencia de género. *Boletín Galego de Medicina Legal e Forense*, 16, Diciembre 2009.
21. *poLCA: Polytomous Variable Latent Class Analysis*. (2006). [Conjunto de datos]. <https://doi.org/10.32614/cran.package.polca>
22. *Preguntas frecuentes: Tipos de violencia contra las mujeres y las niñas*. (2023). ONU Mujeres. <https://www.unwomen.org/es/what-we-do/ending-violence-against-women/faqs/types-of-violence>
23. R Project. (n.d.). Documentation and examples for performing latent class analysis using poLCA and LCA packages. Retrieved from The R Project for Statistical Computing.
24. Schofield, M. R., Maze, M. J., Crump, J. A., Rubach, M. P., Galloway, R., & Sharples, K. J. (2021). On the robustness of latent class models for diagnostic testing with no gold standard. *Statistics in Medicine*, 40(22), 4751-4763. DOI:10.1002/sim.8999
25. Statistical Methods in Medical Research. (2021). Evaluating latent class analysis models: A Monte Carlo simulation study. *Statistical Methods in Medical Research*, 40(22), 4751-4763. DOI:10.1002/sim.8999
26. Tejedor, J. (2019). *Estudio médico-legal de los agresores en la violencia de género*.
27. Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied Latent Class Analysis* (pp. 89-106). Cambridge University Press.
28. Weller, B. E., Bowen, N. K., & Faubert, S. J. (2020). Latent Class Analysis: A Guide to Best Practice. *Social Work Research*, 44(1), 77-88.

ANEXO

```
library(readxl)
library(readxl)
#library(poLCA)
#library(dplyr)
#library(ggplot2)
#library(tidyr)

datos <- read_excel("Datos.xlsx")
datosdf <- data.frame(datos)
View(datosdf)
```

Creo dos filas; una con la suma de cada columna y otra con la suma condicional de cada columna(sólo suma 1 por cada fila) Las divido por 9568 para buscar las prevalencias tradicionales.

```
nombres_columnas <- colnames(datosdf)
sumas_columnas <- colSums(datosdf)/9568
sumas_valores_no_cero <- colSums(datosdf != 0)/9568
sumas <- rbind(nombres_columnas, sumas_columnas, sumas_valores_no_cero)
sumas <- as.data.frame(sumas)
sumas <- rbind(nombres_columnas, as.numeric(as.character(sumas_columnas)), as.numeric(as.character(sumas_valores_no_cero)))
View(sumas)
View(datosdf)
```

Ahora voy con el Análisis de Clases Latentes usando “poLCA”

```
Datos <- read_excel("Datos.xlsx", sheet = "poLCA")
set.seed(1234)
```

1. VIOLENCIA DE CONTROL

```
control <- Datos %>%
  select(C1,C2,C3,C4,C5,C6,C7)
control <- control %>%
  mutate(across(everything(), ~ as.factor(.)))
items <- cbind(C1,C2,C3,C4,C5,C6,C7) ~ 1
```

PARA AJUSTAR MODELOS DE CLASES. Por rentabilidad computacional ajustamos hasta 4 clases

```
mod1 <- poLCA(items, data = control, nclass = 1, maxiter = 1000, nrep = 10)

## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $C1
##           Pr(1) Pr(2)
## class 1: 0.8689 0.1311
##
## $C2
##           Pr(1) Pr(2)
```

```

## class 1: 0.9153 0.0847
##
## $C3
##           Pr(1) Pr(2)
## class 1: 0.8382 0.1618
##
## $C4
##           Pr(1) Pr(2)
## class 1: 0.8502 0.1498
##
## $C5
##           Pr(1) Pr(2)
## class 1: 0.843 0.157
##
## $C6
##           Pr(1) Pr(2)
## class 1: 0.8772 0.1228
##
## $C7
##           Pr(1) Pr(2)
## class 1: 0.911 0.089
##
## Estimated class population shares
## 1
##
## Predicted class memberships (by modal posterior prob.)
## 1
##
## =====
## Fit for 1 latent classes:
## =====
## number of observations: 9568
## number of estimated parameters: 7
## residual degrees of freedom: 120
## maximum log-likelihood: -25362.56
##
## AIC(1): 50739.12
## BIC(1): 50789.28
## G^2(1): 20371.64 (Likelihood ratio/deviance statistic)
## X^2(1): 23851883 (Chi-square goodness of fit)
##

mod2 <- polLCA(items, data = control, nclass = 2, maxiter = 1000, nrep
= 10)

## Model 1: llik = -15960.93 ... best llik = -15960.93
## Model 2: llik = -15960.93 ... best llik = -15960.93
## Model 3: llik = -15960.93 ... best llik = -15960.93
## Model 4: llik = -15960.93 ... best llik = -15960.93
## Model 5: llik = -15960.93 ... best llik = -15960.93
## Model 6: llik = -15960.93 ... best llik = -15960.93
## Model 7: llik = -15960.93 ... best llik = -15960.93
## Model 8: llik = -15960.93 ... best llik = -15960.93
## Model 9: llik = -15960.93 ... best llik = -15960.93

```

```

## Model 10: llik = -15960.93 ... best llik = -15960.93
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $C1
##           Pr(1) Pr(2)
## class 1: 0.9895 0.0105
## class 2: 0.2682 0.7318
##
## $C2
##           Pr(1) Pr(2)
## class 1: 0.9960 0.0040
## class 2: 0.5137 0.4863
##
## $C3
##           Pr(1) Pr(2)
## class 1: 0.9739 0.0261
## class 2: 0.1622 0.8378
##
## $C4
##           Pr(1) Pr(2)
## class 1: 0.9511 0.0489
## class 2: 0.3475 0.6525
##
## $C5
##           Pr(1) Pr(2)
## class 1: 0.9730 0.0270
## class 2: 0.1955 0.8045
##
## $C6
##           Pr(1) Pr(2)
## class 1: 0.9835 0.0165
## class 2: 0.3476 0.6524
##
## $C7
##           Pr(1) Pr(2)
## class 1: 0.9922 0.0078
## class 2: 0.5059 0.4941
##
## Estimated class population shares
## 0.8328 0.1672
##
## Predicted class memberships (by modal posterior prob.)
## 0.8302 0.1698
##
## =====
## Fit for 2 latent classes:
## =====
## number of observations: 9568
## number of estimated parameters: 15
## residual degrees of freedom: 112
## maximum log-likelihood: -15960.93
##
## AIC(2): 31951.86

```

```

## BIC(2): 32059.35
## G^2(2): 1568.384 (Likelihood ratio/deviance statistic)
## X^2(2): 2267.292 (Chi-square goodness of fit)
##

mod3 <- polCA(items, data = control, nclass = 3, maxiter = 1000, nrep
= 10)

## Model 1: llik = -15382.92 ... best llik = -15382.92
## Model 2: llik = -15382.92 ... best llik = -15382.92
## Model 3: llik = -15382.92 ... best llik = -15382.92
## Model 4: llik = -15382.92 ... best llik = -15382.92
## Model 5: llik = -15382.92 ... best llik = -15382.92
## Model 6: llik = -15382.92 ... best llik = -15382.92
## Model 7: llik = -15382.92 ... best llik = -15382.92
## Model 8: llik = -15382.92 ... best llik = -15382.92
## Model 9: llik = -15382.92 ... best llik = -15382.92
## Model 10: llik = -15382.92 ... best llik = -15382.92
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $C1
##           Pr(1) Pr(2)
## class 1:  0.9955 0.0045
## class 2:  0.5947 0.4053
## class 3:  0.0409 0.9591
##
## $C2
##           Pr(1) Pr(2)
## class 1:  0.9986 0.0014
## class 2:  0.8055 0.1945
## class 3:  0.2635 0.7365
##
## $C3
##           Pr(1) Pr(2)
## class 1:  0.9872 0.0128
## class 2:  0.4141 0.5859
## class 3:  0.0165 0.9835
##
## $C4
##           Pr(1) Pr(2)
## class 1:  0.9666 0.0334
## class 2:  0.5208 0.4792
## class 3:  0.2065 0.7935
##
## $C5
##           Pr(1) Pr(2)
## class 1:  0.9841 0.0159
## class 2:  0.4641 0.5359
## class 3:  0.0313 0.9687
##
## $C6
##           Pr(1) Pr(2)
## class 1:  0.9917 0.0083

```

```

## class 2: 0.6190 0.3810
## class 3: 0.1434 0.8566
##
## $C7
##           Pr(1) Pr(2)
## class 1: 0.9956 0.0044
## class 2: 0.8108 0.1892
## class 3: 0.2316 0.7684
##
## Estimated class population shares
## 0.7962 0.1228 0.081
##
## Predicted class memberships (by modal posterior prob.)
## 0.8121 0.106 0.0819
##
## =====
## Fit for 3 latent classes:
## =====
## number of observations: 9568
## number of estimated parameters: 23
## residual degrees of freedom: 104
## maximum log-likelihood: -15382.92
##
## AIC(3): 30811.84
## BIC(3): 30976.67
## G^2(3): 412.3679 (Likelihood ratio/deviance statistic)
## X^2(3): 442.4246 (Chi-square goodness of fit)
##

mod4 <- polCA(items, data = control, nclass = 4, maxiter = 1000, nrep
= 10)

## Model 1: llik = -15379.09 ... best llik = -15379.09
## Model 2: llik = -15296.31 ... best llik = -15296.31
## Model 3: llik = -15298.01 ... best llik = -15296.31
## Model 4: llik = -15296.24 ... best llik = -15296.24
## Model 5: llik = -15296.24 ... best llik = -15296.24
## Model 6: llik = -15322.9 ... best llik = -15296.24
## Model 7: llik = -15298.01 ... best llik = -15296.24
## Model 8: llik = -15296.24 ... best llik = -15296.24
## Model 9: llik = -15296.24 ... best llik = -15296.24
## Model 10: llik = -15344.06 ... best llik = -15296.24
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $C1
##           Pr(1) Pr(2)
## class 1: 0.0470 0.9530
## class 2: 0.9950 0.0050
## class 3: 0.3740 0.6260
## class 4: 0.6818 0.3182
##
## $C2
##           Pr(1) Pr(2)

```

```

## class 1: 0.2549 0.7451
## class 2: 0.9988 0.0012
## class 3: 0.4210 0.5790
## class 4: 0.9835 0.0165
##
## $C3
##           Pr(1) Pr(2)
## class 1: 0.0195 0.9805
## class 2: 0.9859 0.0141
## class 3: 0.5093 0.4907
## class 4: 0.3610 0.6390
##
## $C4
##           Pr(1) Pr(2)
## class 1: 0.2132 0.7868
## class 2: 0.9653 0.0347
## class 3: 0.3826 0.6174
## class 4: 0.5724 0.4276
##
## $C5
##           Pr(1) Pr(2)
## class 1: 0.0282 0.9718
## class 2: 0.9837 0.0163
## class 3: 0.6901 0.3099
## class 4: 0.3515 0.6485
##
## $C6
##           Pr(1) Pr(2)
## class 1: 0.1411 0.8589
## class 2: 0.9914 0.0086
## class 3: 0.8749 0.1251
## class 4: 0.4989 0.5011
##
## $C7
##           Pr(1) Pr(2)
## class 1: 0.2413 0.7587
## class 2: 0.9952 0.0048
## class 3: 0.8194 0.1806
## class 4: 0.8003 0.1997
##
## Estimated class population shares
## 0.0821 0.7999 0.0365 0.0814
##
## Predicted class memberships (by modal posterior prob.)
## 0.0836 0.8121 0.026 0.0783
##
## =====
## Fit for 4 latent classes:
## =====
## number of observations: 9568
## number of estimated parameters: 31
## residual degrees of freedom: 96
## maximum log-likelihood: -15296.24
##

```



```
## AIC(4): 30654.47
## BIC(4): 30876.62
## G^2(4): 238.9964 (Likelihood ratio/deviance statistic)
## X^2(4): 243.0447 (Chi-square goodness of fit)
##
```

Comparo AIC y BIC para elegir el mejor modelo

```
comparacion <- list(mod1, mod2, mod3, mod4)
criterios <- sapply(comparacion, function(x) c(AIC = x$aic, BIC = x$bic))
print(criterios)

##           [,1]      [,2]      [,3]      [,4]
## AIC 50739.12 31951.86 30811.84 30654.47
## BIC 50789.28 32059.35 30976.67 30876.62
```

SELECCIÓN DEL MODELO. Elijo el modelo con 3 clases

```
modelo <- mod3
```

Proporción de individuos en cada clase latente

```
print(modelo$P)

## [1] 0.79616840 0.12279969 0.08103191
```

Prob. de responder si o no a cada ítem, condicionadas a una clase latente

```
probs <- modelo$probs
print(probs)

## $C1
##           Pr(1)      Pr(2)
## class 1: 0.99550090 0.00449907
## class 2: 0.59474263 0.405257369
## class 3: 0.04094131 0.959058690
##
## $C2
##           Pr(1)      Pr(2)
## class 1: 0.9986278 0.001372197
## class 2: 0.8055065 0.194493463
## class 3: 0.2634882 0.736511775
##
## $C3
##           Pr(1)      Pr(2)
## class 1: 0.98724878 0.01275122
## class 2: 0.41413225 0.58586775
## class 3: 0.01652564 0.98347436
##
## $C4
##           Pr(1)      Pr(2)
## class 1: 0.9665641 0.03343592
## class 2: 0.5207696 0.47923040
## class 3: 0.2064842 0.79351577
##
```

```
## $C5
##           Pr(1)      Pr(2)
## class 1: 0.98408259 0.01591741
## class 2: 0.46407220 0.53592780
## class 3: 0.03128406 0.96871594
##
## $C6
##           Pr(1)      Pr(2)
## class 1: 0.9916952 0.008304801
## class 2: 0.6190309 0.380969075
## class 3: 0.1434204 0.856579611
##
## $C7
##           Pr(1)      Pr(2)
## class 1: 0.9955536 0.00444635
## class 2: 0.8107612 0.18923876
## class 3: 0.2315580 0.76844197
```

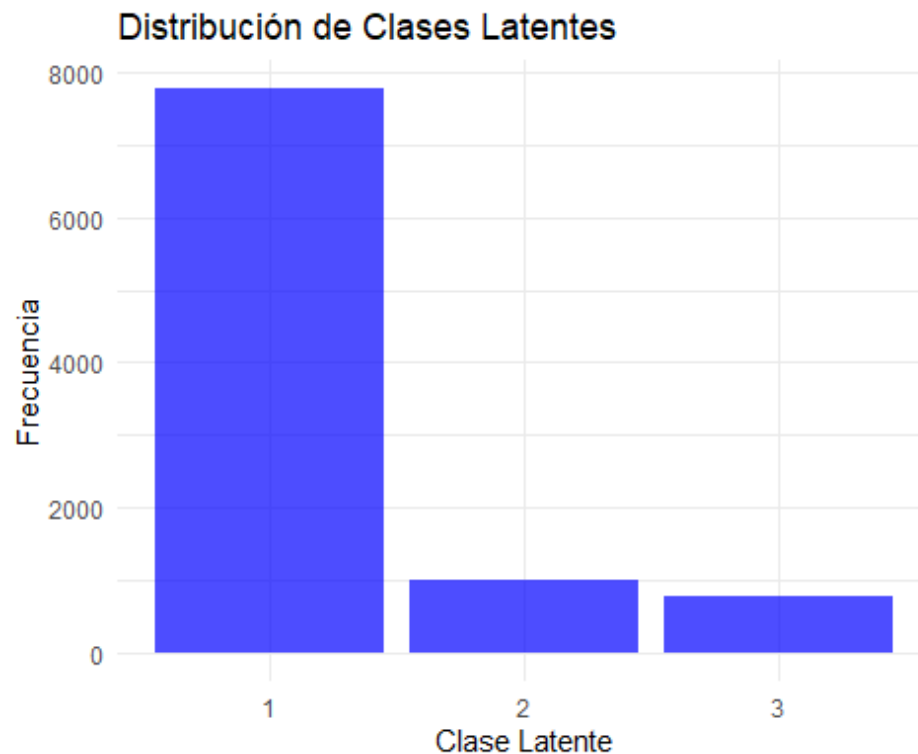
Asignar clases latentes a cada mujer

```
control$class <- apply(modelo$posterior, 1, which.max)
head(control)
```

```
## # A tibble: 6 × 8
##   C1    C2    C3    C4    C5    C6    C7   clase
##   <fct> <fct> <fct> <fct> <fct> <fct> <fct> <int>
## 1 0      0      0      0      0      0      0         1
## 2 0      0      0      0      0      0      0         1
## 3 0      0      0      0      0      0      0         1
## 4 0      0      0      0      0      0      0         1
## 5 0      0      0      0      0      0      0         1
## 6 1      1      1      1      1      1      1         3
```

GRÁFICA DE CLASES LATENTES

```
ggplot(control, aes(x = as.factor(clase))) +
  geom_bar(fill = "blue", alpha = 0.7) +
  labs(title = "Distribución de Clases Latentes",
       x = "Clase Latente",
       y = "Frecuencia") +
  theme_minimal()
```



Gráfica 4: Proporción de mujeres en cada clase latente de Violencia de Control.

Y así para cada violencia.

2. VIOLENCIA ECONOMICA

```
econom <- Datos %>%
  select(E1,E2,E3,E4)
econom <- econom %>%
  mutate(across(everything(), ~ as.factor(.)))
items <- cbind(E1,E2,E3,E4) ~ 1
mod1 <- polCA(items, data = econom, nclass = 1, maxiter = 1000, nrep =
10)

## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $E1
##           Pr(1)  Pr(2)
## class 1: 0.9276 0.0724
##
## $E2
##           Pr(1)  Pr(2)
## class 1: 0.9175 0.0825
##
## $E3
##           Pr(1)  Pr(2)
## class 1: 0.9833 0.0167
```

```

##
## $E4
##           Pr(1) Pr(2)
## class 1:  0.9465 0.0535
##
## Estimated class population shares
## 1
##
## Predicted class memberships (by modal posterior prob.)
## 1
##
## =====
## Fit for 1 latent classes:
## =====
## number of observations: 9568
## number of estimated parameters: 4
## residual degrees of freedom: 11
## maximum log-likelihood: -8021.247
##
## AIC(1): 16050.49
## BIC(1): 16079.16
## G^2(1): 4348.308 (Likelihood ratio/deviance statistic)
## X^2(1): 87890.77 (Chi-square goodness of fit)
##
mod2 <- polCA(items, data = econom, nclass = 2, maxiter = 1000, nrep =
10)

## Model 1: llik = -5866.178 ... best llik = -5866.178
## Model 2: llik = -5866.178 ... best llik = -5866.178
## Model 3: llik = -5866.178 ... best llik = -5866.178
## Model 4: llik = -5866.178 ... best llik = -5866.178
## Model 5: llik = -5866.178 ... best llik = -5866.178
## Model 6: llik = -5866.178 ... best llik = -5866.178
## Model 7: llik = -5866.178 ... best llik = -5866.178
## Model 8: llik = -5866.178 ... best llik = -5866.178
## Model 9: llik = -5866.178 ... best llik = -5866.178
## Model 10: llik = -5866.178 ... best llik = -5866.178
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $E1
##           Pr(1) Pr(2)
## class 1:  0.1693 0.8307
## class 2:  0.9907 0.0093
##
## $E2
##           Pr(1) Pr(2)
## class 1:  0.1202 0.8798
## class 2:  0.9840 0.0160
##
## $E3
##           Pr(1) Pr(2)
## class 1:  0.8476 0.1524

```

```

## class 2: 0.9946 0.0054
##
## $E4
##           Pr(1) Pr(2)
## class 1: 0.4658 0.5342
## class 2: 0.9865 0.0135
##
## Estimated class population shares
## 0.0769 0.9231
##
## Predicted class memberships (by modal posterior prob.)
## 0.0689 0.9311
##
## =====
## Fit for 2 latent classes:
## =====
## number of observations: 9568
## number of estimated parameters: 9
## residual degrees of freedom: 6
## maximum log-likelihood: -5866.178
##
## AIC(2): 11750.36
## BIC(2): 11814.85
## G^2(2): 38.16903 (Likelihood ratio/deviance statistic)
## X^2(2): 61.70317 (Chi-square goodness of fit)
##

mod3 <- polCA(items, data = econom, nclass = 3, maxiter = 1000, nrep =
10)

## Model 1: llik = -5854.62 ... best llik = -5854.62
## Model 2: llik = -5860.517 ... best llik = -5854.62
## Model 3: llik = -5854.656 ... best llik = -5854.62
## Model 4: llik = -5854.621 ... best llik = -5854.62
## Model 5: llik = -5854.627 ... best llik = -5854.62
## Model 6: llik = -5854.62 ... best llik = -5854.62
## Model 7: llik = -5865.554 ... best llik = -5854.62
## Model 8: llik = -5854.622 ... best llik = -5854.62
## Model 9: llik = -5860.95 ... best llik = -5854.62
## Model 10: llik = -5854.714 ... best llik = -5854.62
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $E1
##           Pr(1) Pr(2)
## class 1: 0.9463 0.0537
## class 2: 0.1429 0.8571
## class 3: 0.9924 0.0076
##
## $E2
##           Pr(1) Pr(2)
## class 1: 0.9539 0.0461
## class 2: 0.0846 0.9154
## class 3: 0.9849 0.0151

```

```

##
## $E3
##           Pr(1) Pr(2)
## class 1:  0.9190 0.0810
## class 2:  0.8541 0.1459
## class 3:  1.0000 0.0000
##
## $E4
##           Pr(1) Pr(2)
## class 1:  0.8093 0.1907
## class 2:  0.4606 0.5394
## class 3:  1.0000 0.0000
##
## Estimated class population shares
##  0.0763 0.0722 0.8515
##
## Predicted class memberships (by modal posterior prob.)
##  0.0182 0.0677 0.9141
##
## =====
## Fit for 3 latent classes:
## =====
## number of observations: 9568
## number of estimated parameters: 14
## residual degrees of freedom: 1
## maximum log-likelihood: -5854.62
##
## AIC(3): 11737.24
## BIC(3): 11837.57
## G^2(3): 15.05343 (Likelihood ratio/deviance statistic)
## X^2(3): 14.82005 (Chi-square goodness of fit)
##
## ALERT: iterations finished, MAXIMUM LIKELIHOOD NOT FOUND
##

mod4 <- polCA(items, data = econom, nclass = 4, maxiter = 1000, nrep =
10)

## Model 1: llik = -5854.607 ... best llik = -5854.607
## Model 2: llik = -5854.623 ... best llik = -5854.607
## Model 3: llik = -5854.573 ... best llik = -5854.573
## Model 4: llik = -5854.538 ... best llik = -5854.538
## Model 5: llik = -5853.917 ... best llik = -5853.917
## Model 6: llik = -5847.093 ... best llik = -5847.093
## Model 7: llik = -5854.574 ... best llik = -5847.093
## Model 8: llik = -5853.729 ... best llik = -5847.093
## Model 9: llik = -5847.093 ... best llik = -5847.093
## Model 10: llik = -5854.613 ... best llik = -5847.093
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $E1
##           Pr(1) Pr(2)
## class 1:  0.9072 0.0928

```

```

## class 2: 0.3081 0.6919
## class 3: 0.9950 0.0050
## class 4: 0.0566 0.9434
##
## $E2
##           Pr(1) Pr(2)
## class 1: 0.8975 0.1025
## class 2: 0.0872 0.9128
## class 3: 0.9937 0.0063
## class 4: 0.0620 0.9380
##
## $E3
##           Pr(1) Pr(2)
## class 1: 0.9153 0.0847
## class 2: 0.9898 0.0102
## class 3: 1.0000 0.0000
## class 4: 0.7445 0.2555
##
## $E4
##           Pr(1) Pr(2)
## class 1: 0.8037 0.1963
## class 2: 0.7252 0.2748
## class 3: 1.0000 0.0000
## class 4: 0.2632 0.7368
##
## Estimated class population shares
## 0.0827 0.0377 0.8431 0.0365
##
## Predicted class memberships (by modal posterior prob.)
## 0.0339 0.0483 0.8839 0.034
##
## =====
## Fit for 4 latent classes:
## =====
## number of observations: 9568
## number of estimated parameters: 19
## residual degrees of freedom: -4
## maximum log-likelihood: -5847.093
##
## AIC(4): 11732.19
## BIC(4): 11868.34
## G^2(4): 0.000209064 (Likelihood ratio/deviance statistic)
## X^2(4): 0.0002090559 (Chi-square goodness of fit)
##
## ALERT: iterations finished, MAXIMUM LIKELIHOOD NOT FOUND
##
## ALERT: negative degrees of freedom; respecify model
##

comparacion <- list(mod1, mod2, mod3, mod4)
criterios <- sapply(comparacion, function(x) c(AIC = x$aic, BIC = x$bic))
print(criterios)

```

```
##           [,1]      [,2]      [,3]      [,4]
## AIC 16050.49 11750.36 11737.24 11732.19
## BIC 16079.16 11814.85 11837.57 11868.34
```

Elegimos el modelo con 2 clases

```
modelo <- mod2
```

Proporción de individuos en cada clase latente

```
print(modelo$P)
```

```
## [1] 0.07690481 0.92309519
```

Prob. de responder si o no a cada ítem, condicionadas a una clase latente

```
probs <- modelo$probs
print(probs)
```

```
## $E1
##           Pr(1)      Pr(2)
## class 1: 0.1693208 0.830679242
## class 2: 0.9907423 0.009257661
##
## $E2
##           Pr(1)      Pr(2)
## class 1: 0.1201554 0.87984457
## class 2: 0.9839690 0.01603095
##
## $E3
##           Pr(1)      Pr(2)
## class 1: 0.8475624 0.15243759
## class 2: 0.9945843 0.00541572
##
## $E4
##           Pr(1)      Pr(2)
## class 1: 0.4658124 0.53418763
## class 2: 0.9865343 0.01346568
```

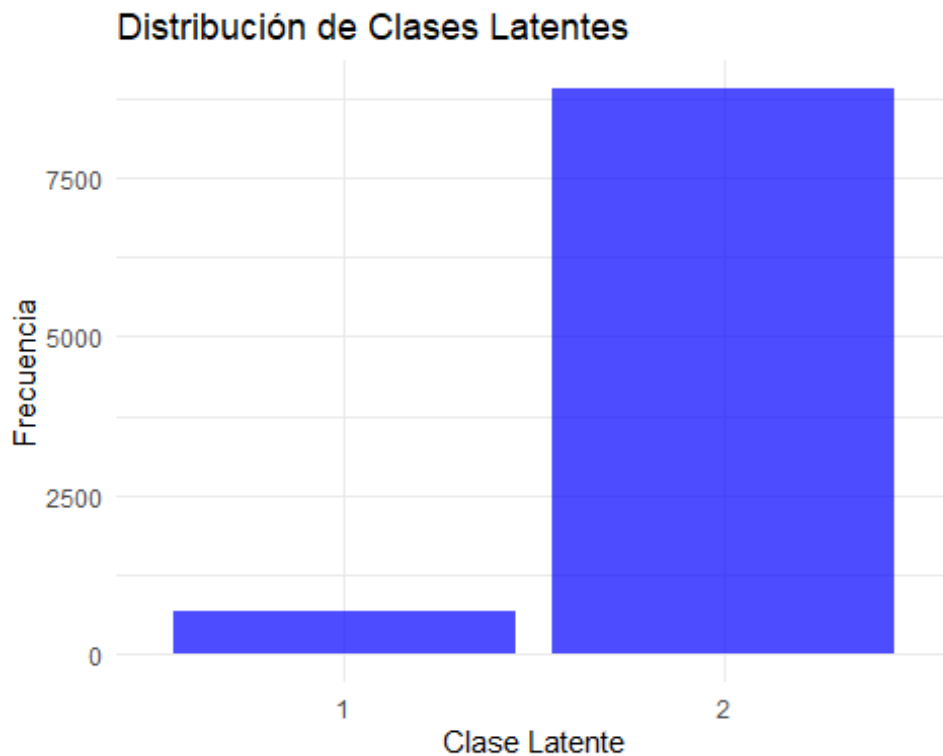
Asignar clases latentes a cada mujer

```
econom$clase <- apply(modelo$posterior, 1, which.max)
head(econom)
```

```
## # A tibble: 6 × 5
##   E1    E2    E3    E4   clase
##   <fct> <fct> <fct> <fct> <int>
## 1 0      0      0      0      2
## 2 0      0      0      0      2
## 3 0      0      0      0      2
## 4 0      0      0      0      2
## 5 0      0      0      0      2
## 6 1      1      0      1      1
```

GRÁFICA DE CLASES LATENTES


```
ggplot(econom, aes(x = as.factor(clase))) +
  geom_bar(fill = "blue", alpha = 0.7) +
  labs(title = "Distribución de Clases Latentes",
       x = "Clase Latente",
       y = "Frecuencia") +
  theme_minimal()
```



Gráfica 5: Proporción de mujeres en cada clase latente de Violencia Económica.

3. VIOLENCIA PSICOLOGICA

```
psico <- Datos %>%
  select(PS1, PS2, PS3, PS4, PS5, PS6, PS7)
psico <- psico %>%
  mutate(across(everything(), ~ as.factor(.)))
items <- cbind(PS1, PS2, PS3, PS4, PS5, PS6, PS7) ~ 1

mod1 <- polCA(items, data = psico, nclass = 1, maxiter = 1000, nrep =
10)

## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $PS1
##           Pr(1) Pr(2)
## class 1: 0.8096 0.1904
##
## $PS2
```

```

##           Pr(1) Pr(2)
## class 1: 0.8587 0.1413
##
## $PS3
##           Pr(1) Pr(2)
## class 1: 0.8722 0.1278
##
## $PS4
##           Pr(1) Pr(2)
## class 1: 0.9082 0.0918
##
## $PS5
##           Pr(1) Pr(2)
## class 1: 0.9554 0.0446
##
## $PS6
##           Pr(1) Pr(2)
## class 1: 0.9378 0.0622
##
## $PS7
##           Pr(1) Pr(2)
## class 1: 0.9615 0.0385
##
## Estimated class population shares
## 1
##
## Predicted class memberships (by modal posterior prob.)
## 1
##
## =====
## Fit for 1 latent classes:
## =====
## number of observations: 9568
## number of estimated parameters: 7
## residual degrees of freedom: 120
## maximum log-likelihood: -20679.69
##
## AIC(1): 41373.38
## BIC(1): 41423.54
## G^2(1): 16289.18 (Likelihood ratio/deviance statistic)
## X^2(1): 40739508 (Chi-square goodness of fit)
##

mod2 <- polCA(items, data = psico, nclass = 2, maxiter = 1000, nrep =
10)

## Model 1: llik = -13167.95 ... best llik = -13167.95
## Model 2: llik = -13167.95 ... best llik = -13167.95
## Model 3: llik = -13167.95 ... best llik = -13167.95
## Model 4: llik = -13167.95 ... best llik = -13167.95
## Model 5: llik = -13167.95 ... best llik = -13167.95
## Model 6: llik = -13167.95 ... best llik = -13167.95
## Model 7: llik = -13167.95 ... best llik = -13167.95
## Model 8: llik = -13167.95 ... best llik = -13167.95

```

```

## Model 9: llik = -13167.95 ... best llik = -13167.95
## Model 10: llik = -13167.95 ... best llik = -13167.95
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $PS1
##           Pr(1) Pr(2)
## class 1: 0.9570 0.0430
## class 2: 0.0597 0.9403
##
## $PS2
##           Pr(1) Pr(2)
## class 1: 0.9846 0.0154
## class 2: 0.2182 0.7818
##
## $PS3
##           Pr(1) Pr(2)
## class 1: 0.9927 0.0073
## class 2: 0.2589 0.7411
##
## $PS4
##           Pr(1) Pr(2)
## class 1: 0.9983 0.0017
## class 2: 0.4499 0.5501
##
## $PS5
##           Pr(1) Pr(2)
## class 1: 0.9945 0.0055
## class 2: 0.7565 0.2435
##
## $PS6
##           Pr(1) Pr(2)
## class 1: 0.9934 0.0066
## class 2: 0.6552 0.3448
##
## $PS7
##           Pr(1) Pr(2)
## class 1: 0.9970 0.0030
## class 2: 0.7813 0.2187
##
## Estimated class population shares
## 0.8357 0.1643
##
## Predicted class memberships (by modal posterior prob.)
## 0.834 0.166
##
## =====
## Fit for 2 latent classes:
## =====
## number of observations: 9568
## number of estimated parameters: 15
## residual degrees of freedom: 112
## maximum log-likelihood: -13167.95
##

```

```

## AIC(2): 26365.91
## BIC(2): 26473.4
## G^2(2): 1265.708 (Likelihood ratio/deviance statistic)
## X^2(2): 2428.436 (Chi-square goodness of fit)
##

mod3 <- polLCA(items, data = psico, nclass = 3, maxiter = 1000, nrep =
10)

## Model 1: llik = -12717.57 ... best llik = -12717.57
## Model 2: llik = -12717.57 ... best llik = -12717.57
## Model 3: llik = -12717.57 ... best llik = -12717.57
## Model 4: llik = -12717.57 ... best llik = -12717.57
## Model 5: llik = -12717.57 ... best llik = -12717.57
## Model 6: llik = -12717.57 ... best llik = -12717.57
## Model 7: llik = -12717.57 ... best llik = -12717.57
## Model 8: llik = -12717.57 ... best llik = -12717.57
## Model 9: llik = -12717.57 ... best llik = -12717.57
## Model 10: llik = -12717.57 ... best llik = -12717.57
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $PS1
##           Pr(1) Pr(2)
## class 1:  0.1345 0.8655
## class 2:  0.0144 0.9856
## class 3:  0.9701 0.0299
##
## $PS2
##           Pr(1) Pr(2)
## class 1:  0.3803 0.6197
## class 2:  0.0884 0.9116
## class 3:  0.9896 0.0104
##
## $PS3
##           Pr(1) Pr(2)
## class 1:  0.4940 0.5060
## class 2:  0.0280 0.9720
## class 3:  0.9951 0.0049
##
## $PS4
##           Pr(1) Pr(2)
## class 1:  0.7606 0.2394
## class 2:  0.0634 0.9366
## class 3:  0.9989 0.0011
##
## $PS5
##           Pr(1) Pr(2)
## class 1:  0.9509 0.0491
## class 2:  0.4903 0.5097
## class 3:  0.9945 0.0055
##
## $PS6
##           Pr(1) Pr(2)

```

```

## class 1: 0.8194 0.1806
## class 2: 0.4524 0.5476
## class 3: 0.9946 0.0054
##
## $PS7
##           Pr(1) Pr(2)
## class 1: 0.9275 0.0725
## class 2: 0.5875 0.4125
## class 3: 0.9973 0.0027
##
## Estimated class population shares
## 0.1148 0.0677 0.8176
##
## Predicted class memberships (by modal posterior prob.)
## 0.0899 0.0763 0.8338
##
## =====
## Fit for 3 latent classes:
## =====
## number of observations: 9568
## number of estimated parameters: 23
## residual degrees of freedom: 104
## maximum log-likelihood: -12717.57
##
## AIC(3): 25481.13
## BIC(3): 25645.96
## G^2(3): 364.9336 (Likelihood ratio/deviance statistic)
## X^2(3): 863.1001 (Chi-square goodness of fit)
##
mod4 <- polCA(items, data = psico, nclass = 4, maxiter = 1000, nrep =
10)

## Model 1: llik = -12669.42 ... best llik = -12669.42
## Model 2: llik = -12644.12 ... best llik = -12644.12
## Model 3: llik = -12643.92 ... best llik = -12643.92
## Model 4: llik = -12669.42 ... best llik = -12643.92
## Model 5: llik = -12669.42 ... best llik = -12643.92
## Model 6: llik = -12669.43 ... best llik = -12643.92
## Model 7: llik = -12648.88 ... best llik = -12643.92
## Model 8: llik = -12644.12 ... best llik = -12643.92
## Model 9: llik = -12669.42 ... best llik = -12643.92
## Model 10: llik = -12644.12 ... best llik = -12643.92
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $PS1
##           Pr(1) Pr(2)
## class 1: 0.0236 0.9764
## class 2: 0.3133 0.6867
## class 3: 0.0148 0.9852
## class 4: 0.9873 0.0127
##
## $PS2

```

```

##           Pr(1)  Pr(2)
## class 1:  0.0893 0.9107
## class 2:  0.5586 0.4414
## class 3:  0.1810 0.8190
## class 4:  0.9975 0.0025
##
## $PS3
##           Pr(1)  Pr(2)
## class 1:  0.0481 0.9519
## class 2:  0.7738 0.2262
## class 3:  0.0596 0.9404
## class 4:  0.9971 0.0029
##
## $PS4
##           Pr(1)  Pr(2)
## class 1:  0.0508 0.9492
## class 2:  0.8982 0.1018
## class 3:  0.3457 0.6543
## class 4:  0.9998 0.0002
##
## $PS5
##           Pr(1)  Pr(2)
## class 1:  0.1191 0.8809
## class 2:  0.9571 0.0429
## class 3:  0.9388 0.0612
## class 4:  0.9954 0.0046
##
## $PS6
##           Pr(1)  Pr(2)
## class 1:  0.3472 0.6528
## class 2:  0.8892 0.1108
## class 3:  0.6691 0.3309
## class 4:  0.9962 0.0038
##
## $PS7
##           Pr(1)  Pr(2)
## class 1:  0.4134 0.5866
## class 2:  0.9511 0.0489
## class 3:  0.8535 0.1465
## class 4:  0.9981 0.0019
##
## Estimated class population shares
##  0.0363 0.1106 0.0701 0.783
##
## Predicted class memberships (by modal posterior prob.)
##  0.0358 0.1073 0.075 0.7818
##
## =====
## Fit for 4 latent classes:
## =====
## number of observations: 9568
## number of estimated parameters: 31
## residual degrees of freedom: 96
## maximum log-likelihood: -12643.92

```

```
##
## AIC(4): 25349.83
## BIC(4): 25571.98
## G^2(4): 217.629 (Likelihood ratio/deviance statistic)
## X^2(4): 504.4827 (Chi-square goodness of fit)
##
## ALERT: iterations finished, MAXIMUM LIKELIHOOD NOT FOUND
##

comparacion <- list(mod1, mod2, mod3, mod4)
criterios <- sapply(comparacion, function(x) c(AIC = x$aic, BIC = x$bic))
print(criterios)

##           [,1]      [,2]      [,3]      [,4]
## AIC 41373.38 26365.91 25481.13 25349.83
## BIC 41423.54 26473.40 25645.96 25571.98
```

SELECCIÓN DEL MODELO. Elegimos el modelo con 3 clases

```
modelo <- mod3
```

Proporción de individuos en cada clase latente

```
print(modelo$P)

## [1] 0.11478804 0.06765414 0.81755782
```

Prob. de responder si o no a cada ítem, condicionadas a una clase latente

```
probs <- modelo$probs
print(probs)

## $PS1
##           Pr(1)      Pr(2)
## class 1: 0.13453865 0.86546135
## class 2: 0.01444089 0.98555911
## class 3: 0.97014932 0.02985068
##
## $PS2
##           Pr(1)      Pr(2)
## class 1: 0.38031290 0.61968710
## class 2: 0.08837803 0.91162197
## class 3: 0.98960724 0.01039276
##
## $PS3
##           Pr(1)      Pr(2)
## class 1: 0.49398943 0.506010572
## class 2: 0.02802743 0.971972568
## class 3: 0.99513187 0.004868127
##
## $PS4
##           Pr(1)      Pr(2)
## class 1: 0.76059000 0.239410000
## class 2: 0.06342797 0.936572031
## class 3: 0.99887490 0.001125098
```

```
##
## $PS5
##           Pr(1)           Pr(2)
## class 1: 0.9508710 0.049128962
## class 2: 0.4902595 0.509740487
## class 3: 0.9944928 0.005507199
##
## $PS6
##           Pr(1)           Pr(2)
## class 1: 0.8194022 0.180597834
## class 2: 0.4524367 0.547563302
## class 3: 0.9946046 0.005395409
##
## $PS7
##           Pr(1)           Pr(2)
## class 1: 0.9274820 0.072517980
## class 2: 0.5875449 0.412455090
## class 3: 0.9972686 0.002731362
```

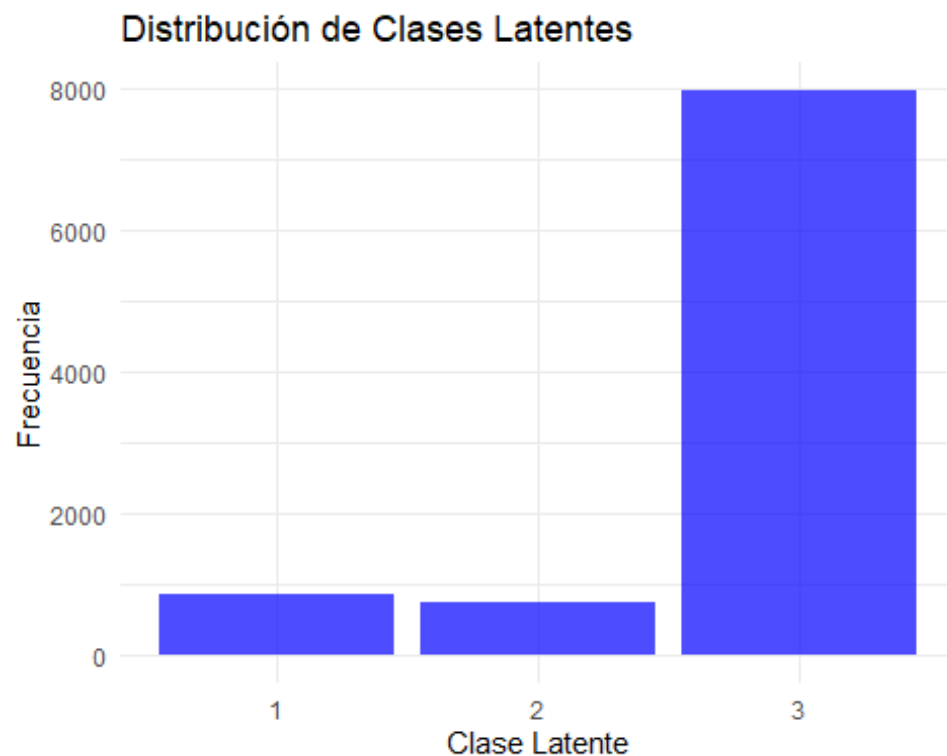
Asignar clases latentes a cada mujer

```
psico$class <- apply(modelo$posterior, 1, which.max)
head(psico)
```

```
## # A tibble: 6 × 8
##   PS1  PS2  PS3  PS4  PS5  PS6  PS7  clase
##   <fct> <fct> <fct> <fct> <fct> <fct> <fct> <int>
## 1 0      0      0      0      0      0      0      3
## 2 0      0      0      0      0      0      0      3
## 3 0      0      0      0      0      0      0      3
## 4 0      0      0      0      0      0      0      3
## 5 0      0      0      0      0      0      0      3
## 6 1      1      1      1      1      1      1      2
```

GRÁFICA DE CLASES LATENTES

```
ggplot(psico, aes(x = as.factor(clase))) +
  geom_bar(fill = "blue", alpha = 0.7) +
  labs(title = "Distribución de Clases Latentes",
       x = "Clase Latente",
       y = "Frecuencia") +
  theme_minimal()
```

Gráfica 6: Proporción de mujeres en cada clase latente de Violencia Psicológica.

4. VIOLENCIA FISICA

```
fisica <- Datos %>%
  select(F1,F2,F3,F4,F5,F6,F7)
fisica <- fisica %>%
  mutate(across(everything(), ~ as.factor(.)))
items <- cbind(F1,F2,F3,F4,F5,F6,F7) ~ 1
```

AJUSTAR MODELOS DE CLASES

```
mod1 <- polCA(items, data = fisica, nclass = 1, maxiter = 1000, nrep =
10)

## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $F1
##           Pr(1)  Pr(2)
## class 1: 0.9296 0.0704
##
## $F2
##           Pr(1)  Pr(2)
## class 1: 0.9129 0.0871
##
## $F3
##           Pr(1)  Pr(2)
## class 1: 0.9504 0.0496
##
```

```

## $F4
##           Pr(1) Pr(2)
## class 1: 0.9614 0.0386
##
## $F5
##           Pr(1) Pr(2)
## class 1: 0.9786 0.0214
##
## $F6
##           Pr(1) Pr(2)
## class 1: 0.9777 0.0223
##
## $F7
##           Pr(1) Pr(2)
## class 1: 0.9576 0.0424
##
## Estimated class population shares
## 1
##
## Predicted class memberships (by modal posterior prob.)
## 1
##
## =====
## Fit for 1 latent classes:
## =====
## number of observations: 9568
## number of estimated parameters: 7
## residual degrees of freedom: 120
## maximum log-likelihood: -12411.12
##
## AIC(1): 24836.24
## BIC(1): 24886.41
## G^2(1): 11569.48 (Likelihood ratio/deviance statistic)
## X^2(1): 3048799474 (Chi-square goodness of fit)
##
mod2 <- polCA(items, data = fisica, nclass = 2, maxiter = 1000, nrep =
10)

## Model 1: llik = -6998 ... best llik = -6998
## Model 2: llik = -6998 ... best llik = -6998
## Model 3: llik = -6998 ... best llik = -6998
## Model 4: llik = -6998 ... best llik = -6998
## Model 5: llik = -6998 ... best llik = -6998
## Model 6: llik = -6998 ... best llik = -6998
## Model 7: llik = -6998 ... best llik = -6998
## Model 8: llik = -6998 ... best llik = -6998
## Model 9: llik = -6998 ... best llik = -6998
## Model 10: llik = -6998 ... best llik = -6998
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $F1
##           Pr(1) Pr(2)

```

```

## class 1: 0.1598 0.8402
## class 2: 0.9909 0.0091
##
## $F2
##           Pr(1) Pr(2)
## class 1: 0.0745 0.9255
## class 2: 0.9797 0.0203
##
## $F3
##           Pr(1) Pr(2)
## class 1: 0.3460 0.6540
## class 2: 0.9985 0.0015
##
## $F4
##           Pr(1) Pr(2)
## class 1: 0.4859 0.5141
## class 2: 0.9993 0.0007
##
## $F5
##           Pr(1) Pr(2)
## class 1: 0.7189 0.2811
## class 2: 0.9993 0.0007
##
## $F6
##           Pr(1) Pr(2)
## class 1: 0.7074 0.2926
## class 2: 0.9993 0.0007
##
## $F7
##           Pr(1) Pr(2)
## class 1: 0.4648 0.5352
## class 2: 0.9968 0.0032
##
## Estimated class population shares
## 0.0738 0.9262
##
## Predicted class memberships (by modal posterior prob.)
## 0.0743 0.9257
##
## =====
## Fit for 2 latent classes:
## =====
## number of observations: 9568
## number of estimated parameters: 15
## residual degrees of freedom: 112
## maximum log-likelihood: -6998
##
## AIC(2): 14026
## BIC(2): 14133.49
## G^2(2): 743.2368 (Likelihood ratio/deviance statistic)
## X^2(2): 1400.193 (Chi-square goodness of fit)
##

```

```

mod3 <- polCA(items, data = fisica, nclass = 3, maxiter = 1000, nrep =
10)

## Model 1: llik = -6744.191 ... best llik = -6744.191
## Model 2: llik = -6744.191 ... best llik = -6744.191
## Model 3: llik = -6779.399 ... best llik = -6744.191
## Model 4: llik = -6744.191 ... best llik = -6744.191
## Model 5: llik = -6744.191 ... best llik = -6744.191
## Model 6: llik = -6744.191 ... best llik = -6744.191
## Model 7: llik = -6744.191 ... best llik = -6744.191
## Model 8: llik = -6744.191 ... best llik = -6744.191
## Model 9: llik = -6744.191 ... best llik = -6744.191
## Model 10: llik = -6744.191 ... best llik = -6744.191
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $F1
##           Pr(1)  Pr(2)
## class 1:  0.3445 0.6555
## class 2:  0.0141 0.9859
## class 3:  0.9928 0.0072
##
## $F2
##           Pr(1)  Pr(2)
## class 1:  0.1657 0.8343
## class 2:  0.0154 0.9846
## class 3:  0.9843 0.0157
##
## $F3
##           Pr(1)  Pr(2)
## class 1:  0.6454 0.3546
## class 2:  0.0215 0.9785
## class 3:  0.9990 0.0010
##
## $F4
##           Pr(1)  Pr(2)
## class 1:  0.8213 0.1787
## class 2:  0.0809 0.9191
## class 3:  0.9995 0.0005
##
## $F5
##           Pr(1)  Pr(2)
## class 1:  0.8953 0.1047
## class 2:  0.5070 0.4930
## class 3:  0.9994 0.0006
##
## $F6
##           Pr(1)  Pr(2)
## class 1:  0.8883 0.1117
## class 2:  0.4912 0.5088
## class 3:  0.9994 0.0006
##
## $F7
##           Pr(1)  Pr(2)

```

```

## class 1: 0.6409 0.3591
## class 2: 0.2899 0.7101
## class 3: 0.9978 0.0022
##
## Estimated class population shares
## 0.0495 0.0318 0.9187
##
## Predicted class memberships (by modal posterior prob.)
## 0.0407 0.0337 0.9257
##
## =====
## Fit for 3 latent classes:
## =====
## number of observations: 9568
## number of estimated parameters: 23
## residual degrees of freedom: 104
## maximum log-likelihood: -6744.191
##
## AIC(3): 13534.38
## BIC(3): 13699.2
## G^2(3): 235.6174 (Likelihood ratio/deviance statistic)
## X^2(3): 346.4002 (Chi-square goodness of fit)
##

mod4 <- polCA(items, data = fisica, nclass = 4, maxiter = 1000, nrep =
10)

## Model 1: llik = -6739.706 ... best llik = -6739.706
## Model 2: llik = -6743.58 ... best llik = -6739.706
## Model 3: llik = -6743.815 ... best llik = -6739.706
## Model 4: llik = -6740.829 ... best llik = -6739.706
## Model 5: llik = -6687.634 ... best llik = -6687.634
## Model 6: llik = -6687.634 ... best llik = -6687.634
## Model 7: llik = -6743.693 ... best llik = -6687.634
## Model 8: llik = -6743.425 ... best llik = -6687.634
## Model 9: llik = -6687.634 ... best llik = -6687.634
## Model 10: llik = -6743.947 ... best llik = -6687.634
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $F1
##           Pr(1) Pr(2)
## class 1: 0.0256 0.9744
## class 2: 0.9934 0.0066
## class 3: 0.0338 0.9662
## class 4: 0.5435 0.4565
##
## $F2
##           Pr(1) Pr(2)
## class 1: 0.0221 0.9779
## class 2: 0.9879 0.0121
## class 3: 0.0464 0.9536
## class 4: 0.2521 0.7479
##

```

```

## $F3
##           Pr(1)  Pr(2)
## class 1:  0.0166 0.9834
## class 2:  0.9989 0.0011
## class 3:  0.1376 0.8624
## class 4:  0.8562 0.1438
##
## $F4
##           Pr(1)  Pr(2)
## class 1:  0.0623 0.9377
## class 2:  0.9995 0.0005
## class 3:  0.3735 0.6265
## class 4:  0.9203 0.0797
##
## $F5
##           Pr(1)  Pr(2)
## class 1:  0.2635 0.7365
## class 2:  0.9996 0.0004
## class 3:  0.8407 0.1593
## class 4:  0.9096 0.0904
##
## $F6
##           Pr(1)  Pr(2)
## class 1:  0.1762 0.8238
## class 2:  0.9996 0.0004
## class 3:  0.8788 0.1212
## class 4:  0.8961 0.1039
##
## $F7
##           Pr(1)  Pr(2)
## class 1:  0.0878 0.9122
## class 2:  0.9986 0.0014
## class 3:  0.5966 0.4034
## class 4:  0.6700 0.3300
##
## Estimated class population shares
##  0.017 0.9119 0.0302 0.0409
##
## Predicted class memberships (by modal posterior prob.)
##  0.0171 0.9257 0.0305 0.0267
##
## =====
## Fit for 4 latent classes:
## =====
## number of observations: 9568
## number of estimated parameters: 31
## residual degrees of freedom: 96
## maximum log-likelihood: -6687.634
##
## AIC(4): 13437.27
## BIC(4): 13659.42
## G^2(4): 122.503 (Likelihood ratio/deviance statistic)
## X^2(4): 183.7021 (Chi-square goodness of fit)
##

```

Comparar AIC y BIC para elegir el mejor modelo

```
comparacion <- list(mod1, mod2, mod3, mod4)
criterios <- sapply(comparacion, function(x) c(AIC = x$aic, BIC = x$bic))
print(criterios)

##           [,1]      [,2]      [,3]      [,4]
## AIC 24836.24 14026.00 13534.38 13437.27
## BIC 24886.41 14133.49 13699.20 13659.42
```

SELECCIÓN DEL MODELO. Elegimos el modelo con 3 clases

```
modelo <- mod3
```

Proporción de individuos en cada clase latente

```
print(modelo$P)

## [1] 0.04949204 0.03181818 0.91868978
```

Prob. de responder si o no a cada ítem, condicionadas a una clase latente

```
probs <- modelo$probs
print(probs)

## $F1
##           Pr(1)      Pr(2)
## class 1: 0.34454625 0.655453751
## class 2: 0.01406896 0.985931041
## class 3: 0.99278007 0.007219926
##
## $F2
##           Pr(1)      Pr(2)
## class 1: 0.16572844 0.83427156
## class 2: 0.01535675 0.98464325
## class 3: 0.98428013 0.01571987
##
## $F3
##           Pr(1)      Pr(2)
## class 1: 0.64535850 0.35464150
## class 2: 0.02145505 0.97854495
## class 3: 0.99895808 0.00104192
##
## $F4
##           Pr(1)      Pr(2)
## class 1: 0.82125315 0.1787468512
## class 2: 0.08090076 0.9190992425
## class 3: 0.99948248 0.0005175245
##
## $F5
##           Pr(1)      Pr(2)
## class 1: 0.8953170 0.1046830356
## class 2: 0.5070162 0.4929837824
## class 3: 0.9993918 0.0006082171
##
```

```
## $F6
##           Pr(1)           Pr(2)
## class 1: 0.8883498 0.1116502396
## class 2: 0.4911574 0.5088425628
## class 3: 0.9994063 0.0005937419
##
## $F7
##           Pr(1)           Pr(2)
## class 1: 0.6408787 0.359121327
## class 2: 0.2899339 0.710066050
## class 3: 0.9977506 0.002249351
```

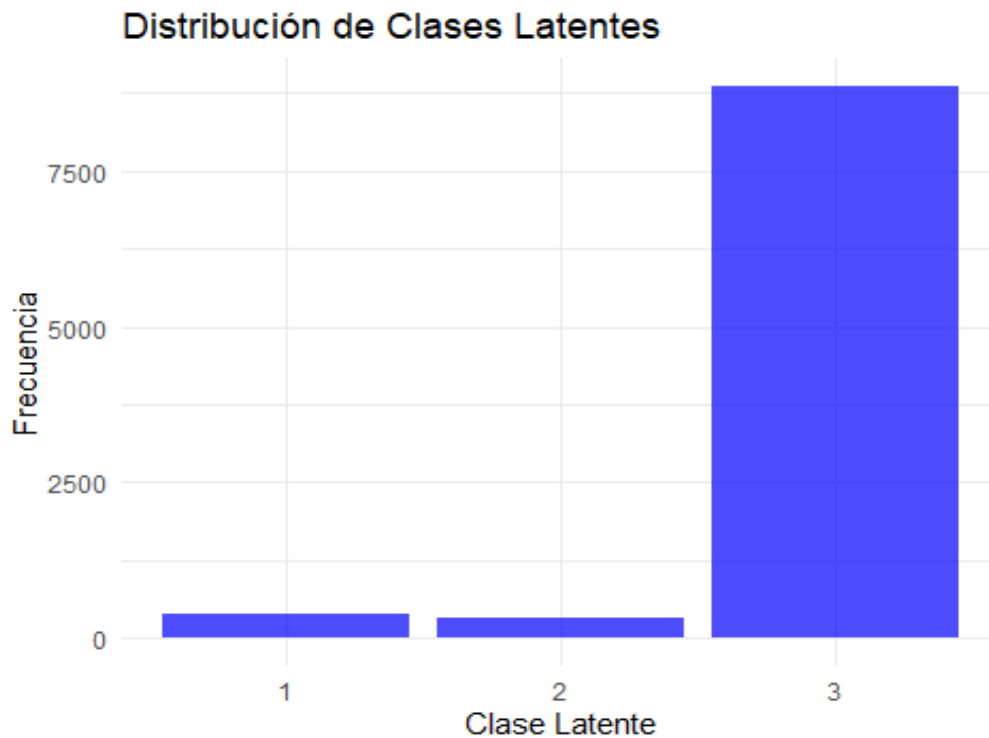
Asignar clases latentes a cada mujer

```
fisica$clase <- apply(modelo$posterior, 1, which.max)
head(fisica)
```

```
## # A tibble: 6 × 8
##   F1     F2     F3     F4     F5     F6     F7     clase
##   <fct> <fct> <fct> <fct> <fct> <fct> <fct> <int>
## 1 0      0      0      0      0      0      0      3
## 2 0      0      0      0      0      0      0      3
## 3 0      0      0      0      0      0      0      3
## 4 0      0      0      0      0      0      0      3
## 5 0      0      0      0      0      0      0      3
## 6 0      1      0      0      0      0      1      1
```

GRÁFICA DE CLASES LATENTES

```
ggplot(fisica, aes(x = as.factor(clase))) +
  geom_bar(fill = "blue", alpha = 0.7) +
  labs(title = "Distribución de Clases Latentes",
       x = "Clase Latente",
       y = "Frecuencia") +
  theme_minimal()
```

Gráfica 7: Proporción de mujeres en cada clase latente de Violencia Física.

5. VIOLENCIA SEXUAL

```
sexual <- Datos %>%
  select(S1,S2,S3,S4,S5,S6,S7,S8)
sexual <- sexual %>%
  mutate(across(everything(), ~ as.factor(.)))
items <- cbind(S1,S2,S3,S4,S5,S6,S7,S8) ~ 1
```

AJUSTAR MODELOS DE CLASES

```
mod1 <- polCA(items, data = sexual, nclass = 1, maxiter = 1000, nrep =
10)

## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $S1
##           Pr(1) Pr(2)
## class 1: 0.9577 0.0423
##
## $S2
##           Pr(1) Pr(2)
## class 1: 0.979 0.021
##
## $S3
##           Pr(1) Pr(2)
## class 1: 0.9474 0.0526
##
```

```

## $S4
##           Pr(1) Pr(2)
## class 1: 0.9359 0.0641
##
## $S5
##           Pr(1) Pr(2)
## class 1: 0.9523 0.0477
##
## $S6
##           Pr(1) Pr(2)
## class 1: 0.945 0.055
##
## $S7
##           Pr(1) Pr(2)
## class 1: 0.959 0.041
##
## $S8
##           Pr(1) Pr(2)
## class 1: 0.9747 0.0253
##
## Estimated class population shares
## 1
##
## Predicted class memberships (by modal posterior prob.)
## 1
##
## =====
## Fit for 1 latent classes:
## =====
## number of observations: 9568
## number of estimated parameters: 8
## residual degrees of freedom: 247
## maximum log-likelihood: -13535.94
##
## AIC(1): 27087.88
## BIC(1): 27145.2
## G^2(1): 14363.18 (Likelihood ratio/deviance statistic)
## X^2(1): 79168532774 (Chi-square goodness of fit)
##

mod2 <- polCA(items, data = sexual, nclass = 2, maxiter = 1000, nrep =
10)

## Model 1: llik = -6848.34 ... best llik = -6848.34
## Model 2: llik = -6848.34 ... best llik = -6848.34
## Model 3: llik = -6848.34 ... best llik = -6848.34
## Model 4: llik = -6848.34 ... best llik = -6848.34
## Model 5: llik = -6848.34 ... best llik = -6848.34
## Model 6: llik = -6848.34 ... best llik = -6848.34
## Model 7: llik = -6848.34 ... best llik = -6848.34
## Model 8: llik = -6848.34 ... best llik = -6848.34
## Model 9: llik = -6848.34 ... best llik = -6848.34
## Model 10: llik = -6848.34 ... best llik = -6848.34
## Conditional item response (column) probabilities,

```

```

## by outcome variable, for each class (row)
##
## $S1
##           Pr(1) Pr(2)
## class 1: 0.9985 0.0015
## class 2: 0.3809 0.6191
##
## $S2
##           Pr(1) Pr(2)
## class 1: 0.9981 0.0019
## class 2: 0.7090 0.2910
##
## $S3
##           Pr(1) Pr(2)
## class 1: 0.9959 0.0041
## class 2: 0.2630 0.7370
##
## $S4
##           Pr(1) Pr(2)
## class 1: 0.9924 0.0076
## class 2: 0.1380 0.8620
##
## $S5
##           Pr(1) Pr(2)
## class 1: 0.9947 0.0053
## class 2: 0.3533 0.6467
##
## $S6
##           Pr(1) Pr(2)
## class 1: 0.9948 0.0052
## class 2: 0.2419 0.7581
##
## $S7
##           Pr(1) Pr(2)
## class 1: 0.9990 0.0010
## class 2: 0.3937 0.6063
##
## $S8
##           Pr(1) Pr(2)
## class 1: 0.9995 0.0005
## class 2: 0.6242 0.3758
##
## Estimated class population shares
## 0.9339 0.0661
##
## Predicted class memberships (by modal posterior prob.)
## 0.9325 0.0675
##
## =====
## Fit for 2 latent classes:
## =====
## number of observations: 9568
## number of estimated parameters: 17
## residual degrees of freedom: 238

```

```

## maximum log-likelihood: -6848.34
##
## AIC(2): 13730.68
## BIC(2): 13852.51
## G^2(2): 987.9839 (Likelihood ratio/deviance statistic)
## X^2(2): 1720.297 (Chi-square goodness of fit)
##

mod3 <- polCA(items, data = sexual, nclass = 3, maxiter = 1000, nrep =
10)

## Model 1: llik = -6550.256 ... best llik = -6550.256
## Model 2: llik = -6550.256 ... best llik = -6550.256
## Model 3: llik = -6550.256 ... best llik = -6550.256
## Model 4: llik = -6550.256 ... best llik = -6550.256
## Model 5: llik = -6550.256 ... best llik = -6550.256
## Model 6: llik = -6550.256 ... best llik = -6550.256
## Model 7: llik = -6550.256 ... best llik = -6550.256
## Model 8: llik = -6550.256 ... best llik = -6550.256
## Model 9: llik = -6550.256 ... best llik = -6550.256
## Model 10: llik = -6550.256 ... best llik = -6550.256
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $S1
##           Pr(1) Pr(2)
## class 1:  0.1183 0.8817
## class 2:  0.9989 0.0011
## class 3:  0.6190 0.3810
##
## $S2
##           Pr(1) Pr(2)
## class 1:  0.5904 0.4096
## class 2:  0.9985 0.0015
## class 3:  0.8132 0.1868
##
## $S3
##           Pr(1) Pr(2)
## class 1:  0.0223 0.9777
## class 2:  0.9968 0.0032
## class 3:  0.4874 0.5126
##
## $S4
##           Pr(1) Pr(2)
## class 1:  0.0115 0.9885
## class 2:  0.9943 0.0057
## class 3:  0.2756 0.7244
##
## $S5
##           Pr(1) Pr(2)
## class 1:  0.2014 0.7986
## class 2:  0.9958 0.0042
## class 3:  0.5026 0.4974
##

```

```

## $S6
##           Pr(1) Pr(2)
## class 1:  0.0052 0.9948
## class 2:  0.9957 0.0043
## class 3:  0.4658 0.5342
##
## $S7
##           Pr(1) Pr(2)
## class 1:  0.0233 0.9767
## class 2:  0.9993 0.0007
## class 3:  0.7085 0.2915
##
## $S8
##           Pr(1) Pr(2)
## class 1:  0.3276 0.6724
## class 2:  0.9996 0.0004
## class 3:  0.8663 0.1337
##
## Estimated class population shares
##  0.0287 0.9294 0.0419
##
## Predicted class memberships (by modal posterior prob.)
##  0.0287 0.9325 0.0388
##
## =====
## Fit for 3 latent classes:
## =====
## number of observations: 9568
## number of estimated parameters: 26
## residual degrees of freedom: 229
## maximum log-likelihood: -6550.256
##
## AIC(3): 13152.51
## BIC(3): 13338.83
## G^2(3): 391.8148 (Likelihood ratio/deviance statistic)
## X^2(3): 453.071 (Chi-square goodness of fit)
##
mod4 <- polCA(items, data = sexual, nclass = 4, maxiter = 1000, nrep =
10)

## Model 1: llik = -6549.546 ... best llik = -6549.546
## Model 2: llik = -6549.809 ... best llik = -6549.546
## Model 3: llik = -6508.762 ... best llik = -6508.762
## Model 4: llik = -6510.311 ... best llik = -6508.762
## Model 5: llik = -6546.266 ... best llik = -6508.762
## Model 6: llik = -6510.306 ... best llik = -6508.762
## Model 7: llik = -6521.284 ... best llik = -6508.762
## Model 8: llik = -6510.304 ... best llik = -6508.762
## Model 9: llik = -6510.307 ... best llik = -6508.762
## Model 10: llik = -6524.363 ... best llik = -6508.762
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##

```

```

## $S1
##           Pr(1)  Pr(2)
## class 1:  0.7827 0.2173
## class 2:  0.9989 0.0011
## class 3:  0.5812 0.4188
## class 4:  0.1298 0.8702
##
## $S2
##           Pr(1)  Pr(2)
## class 1:  0.8931 0.1069
## class 2:  0.9985 0.0015
## class 3:  0.7886 0.2114
## class 4:  0.6000 0.4000
##
## $S3
##           Pr(1)  Pr(2)
## class 1:  0.7904 0.2096
## class 2:  0.9968 0.0032
## class 3:  0.4094 0.5906
## class 4:  0.0174 0.9826
##
## $S4
##           Pr(1)  Pr(2)
## class 1:  0.4625 0.5375
## class 2:  0.9944 0.0056
## class 3:  0.2108 0.7892
## class 4:  0.0161 0.9839
##
## $S5
##           Pr(1)  Pr(2)
## class 1:  0.4314 0.5686
## class 2:  0.9957 0.0043
## class 3:  0.5272 0.4728
## class 4:  0.2244 0.7756
##
## $S6
##           Pr(1)  Pr(2)
## class 1:  0.1207 0.8793
## class 2:  0.9956 0.0044
## class 3:  0.6150 0.3850
## class 4:  0.0186 0.9814
##
## $S7
##           Pr(1)  Pr(2)
## class 1:  0.2277 0.7723
## class 2:  0.9992 0.0008
## class 3:  0.9272 0.0728
## class 4:  0.0353 0.9647
##
## $S8
##           Pr(1)  Pr(2)
## class 1:  0.7742 0.2258
## class 2:  0.9996 0.0004
## class 3:  0.9053 0.0947

```

```
## class 4: 0.3613 0.6387
##
## Estimated class population shares
## 0.0108 0.9296 0.0287 0.0309
##
## Predicted class memberships (by modal posterior prob.)
## 0.0094 0.9325 0.0272 0.0309
##
## =====
## Fit for 4 latent classes:
## =====
## number of observations: 9568
## number of estimated parameters: 35
## residual degrees of freedom: 220
## maximum log-likelihood: -6508.762
##
## AIC(4): 13087.52
## BIC(4): 13338.34
## G^2(4): 308.8266 (Likelihood ratio/deviance statistic)
## X^2(4): 388.3866 (Chi-square goodness of fit)
##
```

Comparar AIC y BIC para elegir el mejor modelo

```
comparacion <- list(mod1, mod2, mod3, mod4)
criterios <- sapply(comparacion, function(x) c(AIC = x$aic, BIC = x$bic))
print(criterios)

##           [,1]      [,2]      [,3]      [,4]
## AIC 27087.88 13730.68 13152.51 13087.52
## BIC 27145.20 13852.51 13338.83 13338.34
```

SELECCIÓN DEL MODELO y el número de clases latentes

```
modelo <- mod3
```

Proporción de individuos en cada clase latente

```
print(modelo$P)

## [1] 0.02874587 0.92937743 0.04187669
```

Prob. de responder si o no a cada ítem, condicionadas a una clase latente

```
probs <- modelo$probs
print(probs)

## $S1
##           Pr(1)      Pr(2)
## class 1: 0.1182546 0.881745350
## class 2: 0.9988944 0.001105561
## class 3: 0.6190107 0.380989302
##
## $S2
##           Pr(1)      Pr(2)
```

```
## class 1: 0.5904485 0.409551452
## class 2: 0.9984819 0.001518091
## class 3: 0.8131722 0.186827824
##
## $S3
##           Pr(1)           Pr(2)
## class 1: 0.02225677 0.977743231
## class 2: 0.99677419 0.003225807
## class 3: 0.48737606 0.512623943
##
## $S4
##           Pr(1)           Pr(2)
## class 1: 0.01152017 0.988479833
## class 2: 0.99427855 0.005721447
## class 3: 0.27559642 0.724403577
##
## $S5
##           Pr(1)           Pr(2)
## class 1: 0.2013770 0.798622979
## class 2: 0.9958323 0.004167666
## class 3: 0.5026252 0.497374842
##
## $S6
##           Pr(1)           Pr(2)
## class 1: 0.005194207 0.994805793
## class 2: 0.995689672 0.004310328
## class 3: 0.465754505 0.534245495
##
## $S7
##           Pr(1)           Pr(2)
## class 1: 0.02333704 0.97666296
## class 2: 0.99926201 0.00073799
## class 3: 0.70845359 0.29154641
##
## $S8
##           Pr(1)           Pr(2)
## class 1: 0.3276134 0.6723866051
## class 2: 0.9996075 0.0003925139
## class 3: 0.8662858 0.1337142105
```

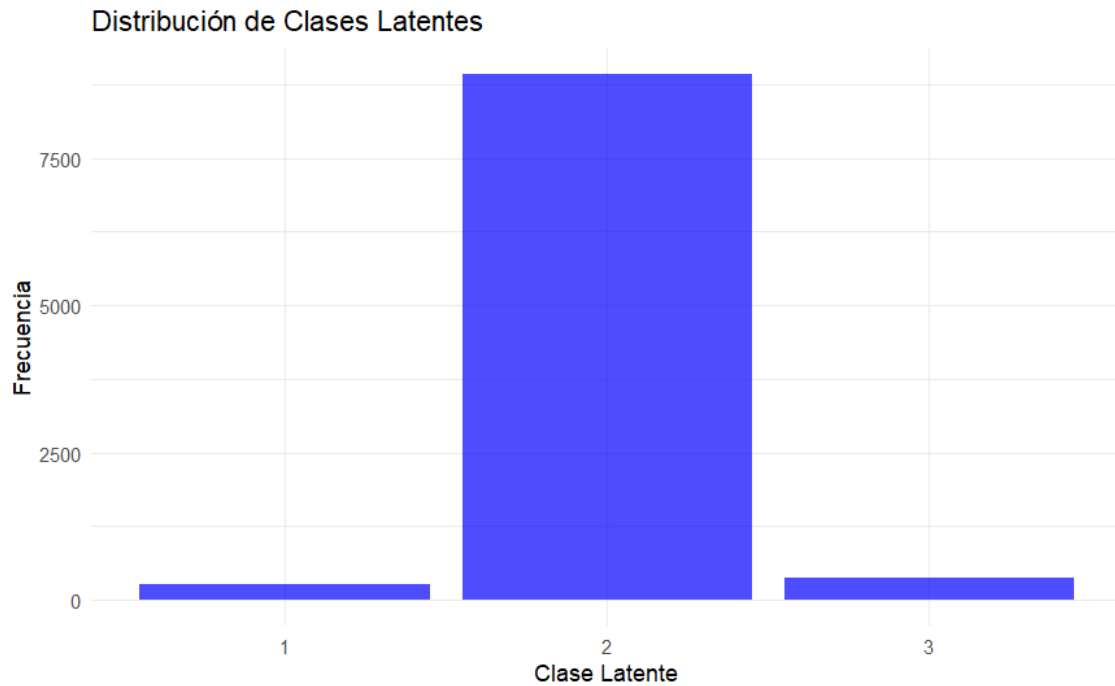
Asigna clases latentes a cada mujer

```
sexual$clase <- apply(modelo$posterior, 1, which.max)
head(sexual)
```

```
## # A tibble: 6 × 9
##   S1    S2    S3    S4    S5    S6    S7    S8    clase
##   <fct> <fct> <fct> <fct> <fct> <fct> <fct> <fct> <int>
## 1 0      0      0      0      0      0      0      0      2
## 2 0      0      0      0      0      0      0      0      2
## 3 0      0      0      0      0      0      0      0      2
## 4 0      0      0      0      0      0      0      0      2
## 5 0      0      0      0      0      0      0      0      2
## 6 0      0      0      0      0      0      0      0      2
```


GRÁFICA DE CLASES LATENTES

```
ggplot(sexual, aes(x = as.factor(clase))) +  
  geom_bar(fill = "blue", alpha = 0.7) +  
  labs(title = "Distribución de Clases Latentes",  
        x = "Clase Latente",  
        y = "Frecuencia") +  
  theme_minimal()
```



Gráfica 8: Proporción de mujeres en cada clase latente de Violencia Sexual.