

Proyecto de Reconocimiento Estadístico de Patrones: Análisis del beneficio de los hombres en la vida doméstica a escala global y regional

Juan Luis Baldelomar Cabrera y Andrea Quintanilla Carranza

8 de junio de 2021

0. Intrucciones. Proyecto de análisis de datos

El resultado debe ser una narrativa visual que se debe presentar de manera electrónica (como si fuera para un concurso de datos). Se trabaja en equipos de dos. Cada equipo tiene toda la libertad de elegir las preguntas de interés a las cuales se quieren enfocar (sugerencia: conviene acotar las preguntas de interés e ir por profundidad para que el resultado tenga más coherencia). La idea es por supuesto aprovechar lo más posible las herramientas que vimos en clase.

Si aplica, enfócate a preguntas ligadas a México o de relevancia especial para ti (para nosotros). La mitad de la calificación es para evaluar los aspectos técnicos y la otra mitad es para la presentación / la narrativa.

I. Descripción general de los datos y de las preguntas de interés

En este trabajo ocupamos el conjunto de datos que se encuentran en el archivo `sog_agg_country.xlsx` obtenidos del proyecto en <https://www.equalityathome.org/>. Es una encuesta realizada en colaboración por varias instituciones: Facebook, el Banco Mundial, UN Women, Equal MEasures 2030 y Ladysmith. Ésta fue aplicada en julio del 2020, a aproximadamente medio millón de personas en 208 países, con la finalidad de capturar dinámicas en el hogar ocasionadas por la pandemia del COVID-19.

La encuesta contiene preguntas respecto a actitudes y creencias en relación al género. Además de preguntas en relación a la participación de las personas en labores domésticos y de cuidado no remunerados. Específicamente, las preguntas se dividen en cuatro categorías: antecedentes, capacidad financiera y de decisión, trabajo de cuidado no remunerado e información demográfica. Es importante resaltar que esta encuesta está limitada a participantes con acceso a internet, en particular, a miembros de Facebook, por lo cual no representa en una totalidad a toda la población de un país. Además, debido a que la cantidad de respuestas de personas que poseen un género no binario es demasiado baja, se incluyen únicamente respuestas de personas que se identifican como mujer u hombre.

La pregunta guía que tratamos de responder es si la participación de los hombres en tareas domésticas es menor en relación con la participación de las mujeres. Este análisis se desea llevar a cabo a escala global y regional. Podemos adelantar que en efecto la participación de los hombres en tareas domésticas es menor en relación a la participación por parte de las mujeres, por lo tanto se desea comprender las relaciones entre algunas de las preguntas realizadas con este factor. Por ejemplo, es de interés analizar si estas diferencias del desempeño de actividades domésticas tiene alguna relación con creencias culturales o prejuicios, con trabajo remunerado o con poder adquisitivo y de decisión.

Las observaciones corresponden a los promedios de las respuestas por cada país y por género y un promedio de ambos géneros. Para realizar el siguiente trabajo, se decidió considerar sólo las observaciones de la forma país-género, con género femenino (Gender=1) o masculino (Gender=0). Es decir, se omitieron las observaciones correspondientes a promedios entre ambos géneros. En la siguiente figura se muestra una descripción breve del balance de los datos con los que se trabajaron entre regiones, número de países y entre género.

center

89 Variables

246 Observaciones

Región

| | | |
|-----|---------|----------|
| n | missing | distinct |
| 246 | 0 | 7 |

East Asia Pacific (24, 0.098), Europe and Central Asia (78, 0.317),
 Latin America and Caribbean (42, 0.171), Middle East and North Africa (34, 0.138),
 North America (4, 0.016), South Asia (14, 0.057), Sub-Saharan Africa (50, 0.203)

País

| | | |
|-----|---------|----------|
| n | missing | distinct |
| 246 | 0 | 123 |

Género

| | | | | | | |
|-----|---------|----------|------|-----|------|-------|
| n | missing | distinct | Info | Sum | Mean | Gmd |
| 246 | 0 | 2 | 0.75 | 123 | 0.5 | 0.502 |

II. Exploración de los datos: ¿Hay un desbalance en el trabajo no remunerado?

Preguntas/variables con mayor contraste

Como un primer acercamiento, ya que estamos interesados en las diferencias entre las respuestas de hombres y mujeres, se dividió a los datos de los países según su género y se calcularon las diferencias entre géneros, de los porcentajes de respuesta por cada país. Después, se calculó el promedio de esas diferencias por pregunta y se seleccionaron siete de las preguntas que más mujeres respondieron positivamente, y el mismo número de preguntas en donde un porcentaje mayor de hombres respondió positivamente. (Se eligieron siete pues hasta ese número había una diferencia que aún no se acercaba al promedio global de diferencias).

En las siguientes figuras se muestra el porcentaje de personas de cada país que respondieron a la pregunta positivamente, dividida por género. Las líneas verticales muestran la densidad de datos

en cada región. En la primera se agrupan las siete correspondientes a las mujeres y en la segunda a las de los hombres.

Figura 1: Cada gráfica corresponde a una pregunta: porcentaje de respuestas afirmativas vs. género. Agrupación por género: mujeres (morado), verde (hombres). Porcentajes más altos en mujeres.

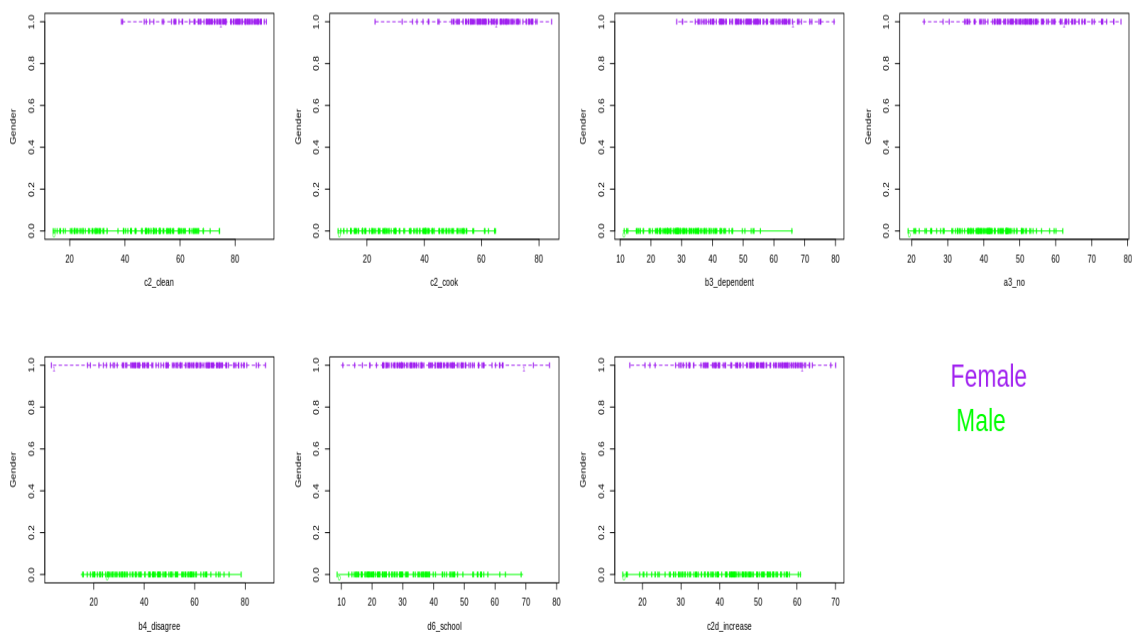
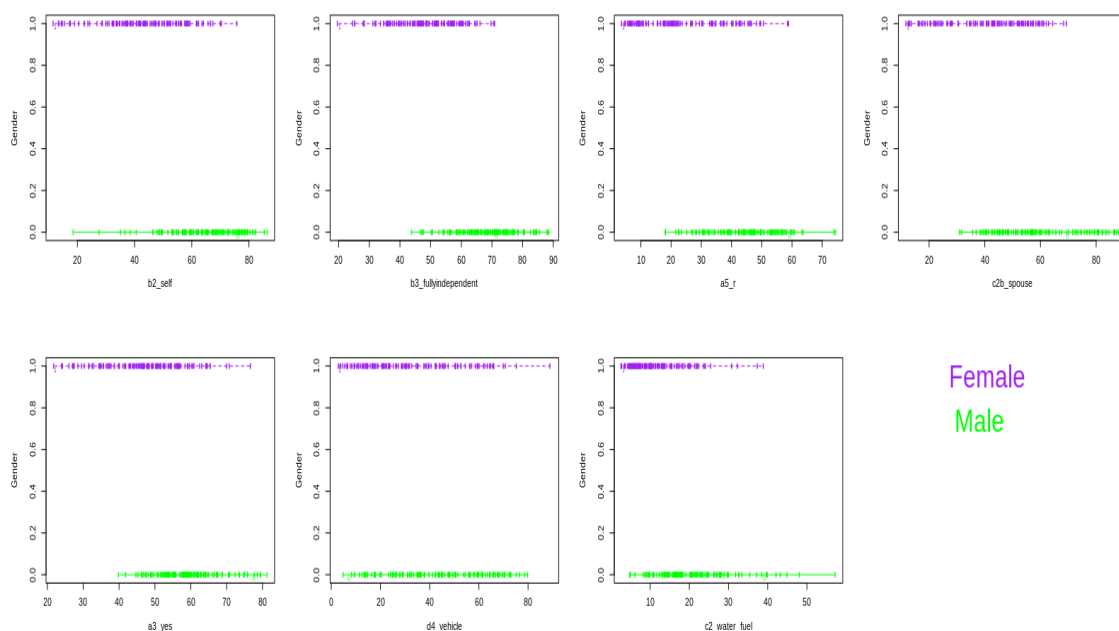


Figura 2: Cada gráfica corresponde a una pregunta: porcentaje de respuestas afirmativas vs. género. Agrupación por género: mujeres (morado), verde (hombres). Porcentajes más altos en hombres.



Las preguntas de la figura 1, con niveles de respuestas afirmativos más altos en mujeres respecto a hombres, corresponden a:

- En circunstancias normales, ¿te haces cargo de limpiar tu casa al menos la mitad de las veces?
- En circunstancias normales, ¿te haces cargo de cocinar en tu casa al menos la mitad de las veces?
- ¿Durante el último año has sido predominantemente dependiente financieramente?
- ¿Durante la última semana no has hecho ni un sólo trabajo que sea remunerado económicamente?
- ¿Estás en desacuerdo con la frase: “Los hombres son los responsables de encargarse de los gastos domésticos, incluso cuando su esposa los pueda apoyar”?
- Debido a la pandemia ¿tus actividades escolares se han visto reducidas o canceladas?
- Durante la pandemia, ¿el tiempo que has dedicado al cuidado de miembros de la familia ha aumentado?

Mientras que en la figura 2, con niveles de respuestas afirmativos más altos en hombres, las preguntas son:

- ¿Eres el que aporta mayor dinero a los gastos domésticos en tu casa?

- ¿Durante el último año has sido predominantemente totalmente independiente financieramente?
- ¿Eres el responsable de tomar las decisiones y aportar dinero a tu hogar?
- ¿Tu pareja normalmente se encarga del cuidado de los miembros de la familia?
- ¿Durante la última semana has hecho algún trabajo que sea remunerado económicamente?
- ¿Eres dueño de algún vehículo?
- En circunstancias normales, ¿te haces cargo de obtener agua o combustible en tu casa al menos la mitad de las veces?

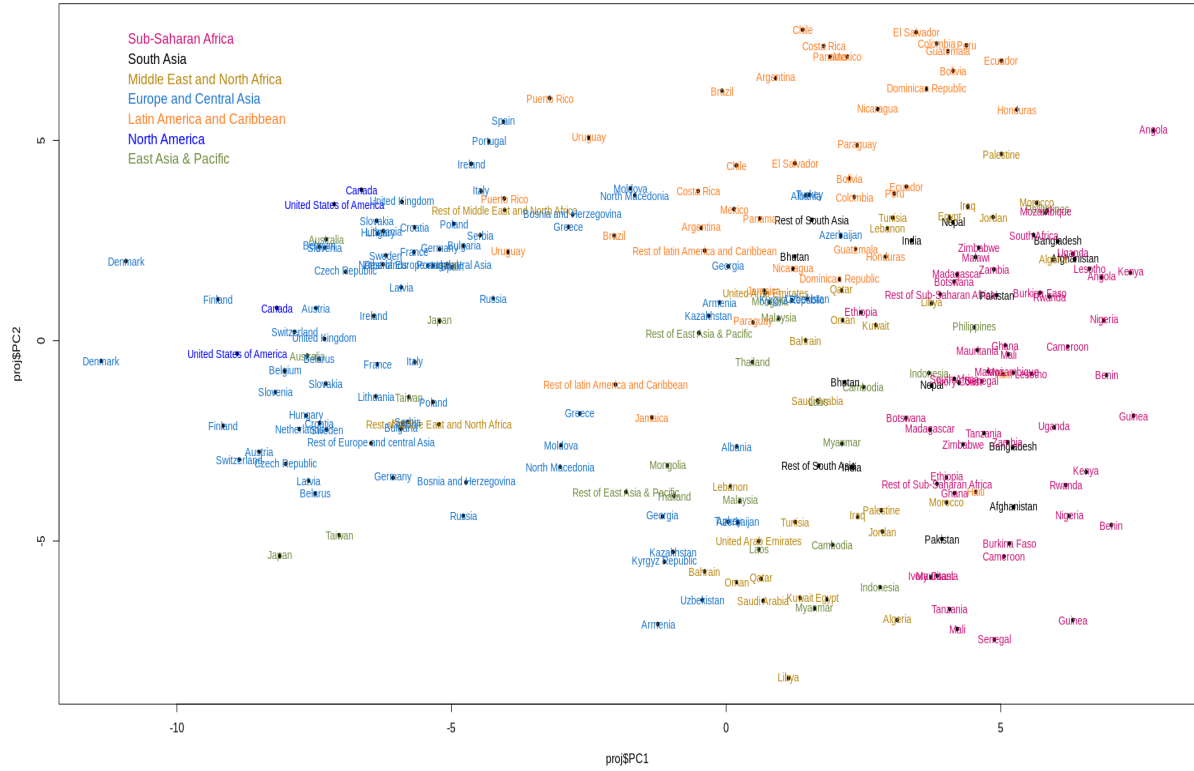
Observamos que, salvo en tres casos, todas las preguntas versan sobre distribución de trabajo. Y en todos esos casos, los trabajos predominantemente realizados por mujeres son no remunerados y los hechos por hombres sí lo son.

Similitudes en las Observaciones

PCA escala global

Para buscar similitudes en los datos podemos realizar un agrupamiento de los datos. Realizar agrupamiento de los datos puede ser de mucha utilidad para revelar características en común que comparten los elementos u observaciones que pertenecen a un mismo grupo ya que no es necesario supervisar los datos, es decir no es necesario indicarle a los algoritmos bajo que características deben realizar los agrupamientos. Como los datos fueron tomados a escala global, las variables económicas y políticas varían ampliamente por región; a nosotros nos interesan las diferencias respecto a género, por lo tanto, se decidió estudiar agrupamientos tomando en cuenta las observaciones que pertenecen a una misma región. En la figura 3, tras realizar una proyección en las primeras dos componentes principales, se puede observar que, efectivamente, existe mayor homogeneidad por región. El PCA mostró que con las primeras dos componentes se explica alrededor del 42 % de la varianza.

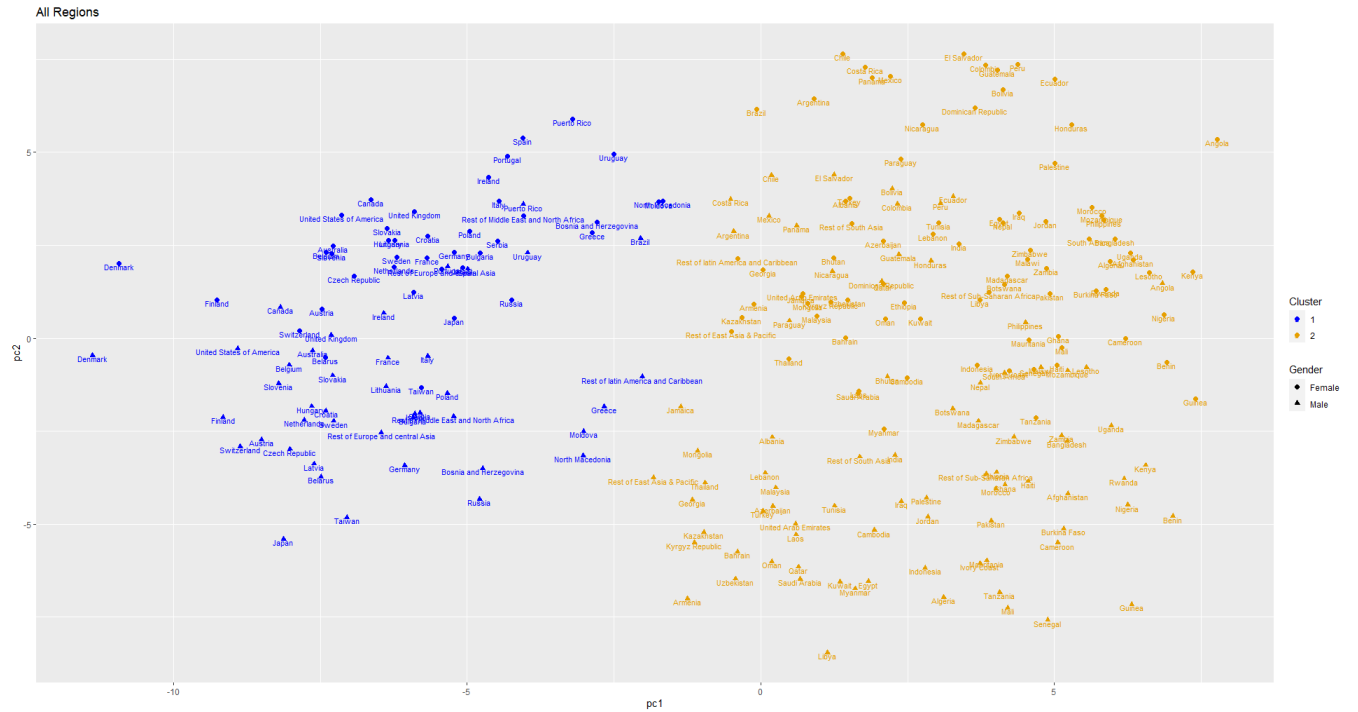
Figura 3: Proyección en las primeras dos CP. Colores señalan región, los nombres de los países se observan en texto.



Antes de llevar a cabo el agrupamiento, se normalizaron los datos debido a las diferentes escalas que se manejaban en las respuestas. Para agrupar los datos se utilizó el algoritmo de K-Medias y posteriormente se visualizaron realizando una proyección de estos a \mathbb{R}^2 a través de los primeros dos componentes que se obtienen con PCA.

En la figura 4 podemos observar el agrupamiento a escala global de todas las observaciones para $k = 2$. Los clusters se encuentran señalados por diferentes colores; el género femenino se identifica por un punto y el masculino por un triángulo. Es difícil determinar un patrón a simple vista de las diferencias entre los géneros. Si aumentamos el número de clusters a 3 o 4 tampoco se logra observar un patrón claro, y más parece suceder que los datos quedan agrupados por la región a la que pertenecen. Esto motiva a delimitar los datos y realizar el agrupamiento de las observaciones pertenecientes a una región en particular.

Figura 4: Agrupamiento a escala global



Sin embargo, es interesante observar los pesos de las preguntas asociados a las componentes principales en las figuras 5 y 6. En ellas se colorearon las preguntas, según nuestras preguntas de interés. Es decir, ¿cómo se relacionaría el trabajo en el hogar con rasgos como: opiniones sobre equidad, características económicas, demográficas, con COVID?

Figura 5: Pesos de primera componente principal. Tipo de pregunta señalada por color.

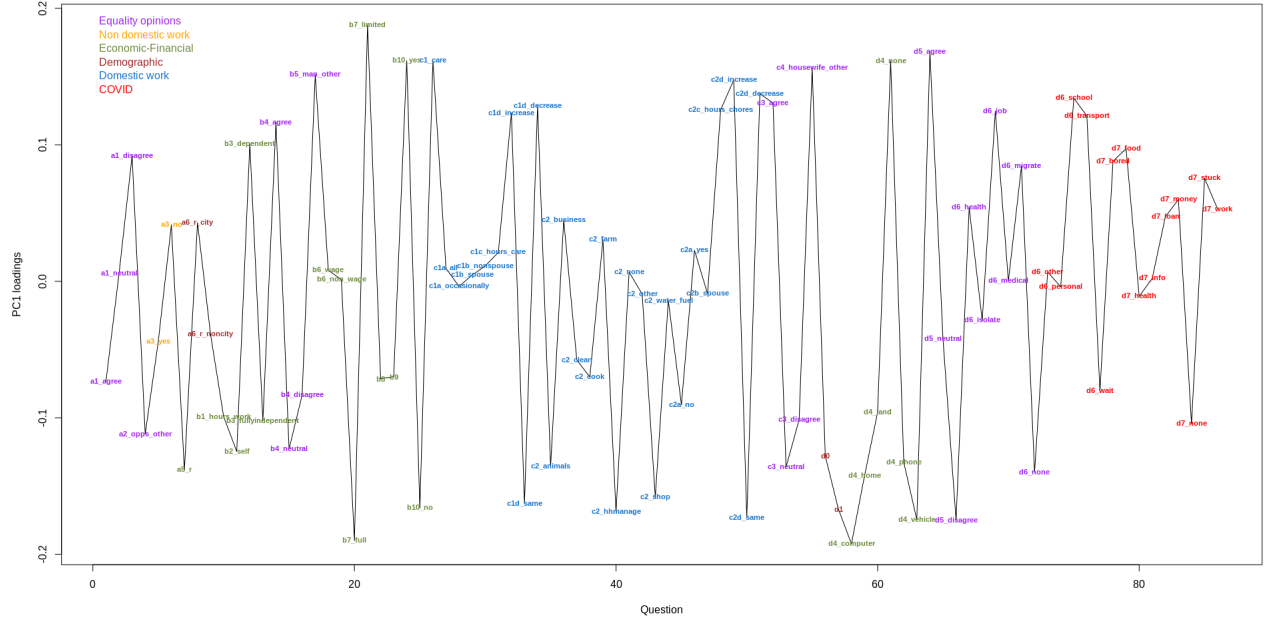
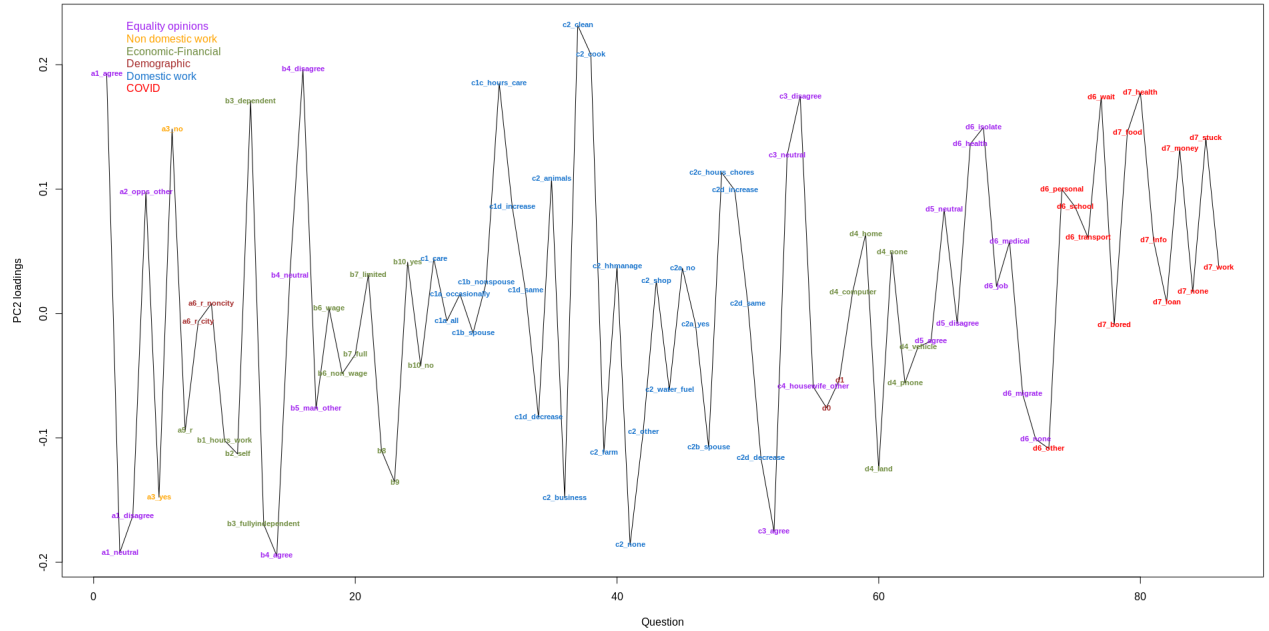


Figura 6: Pesos de segunda componente principal. Tipo de pregunta señalada por color.



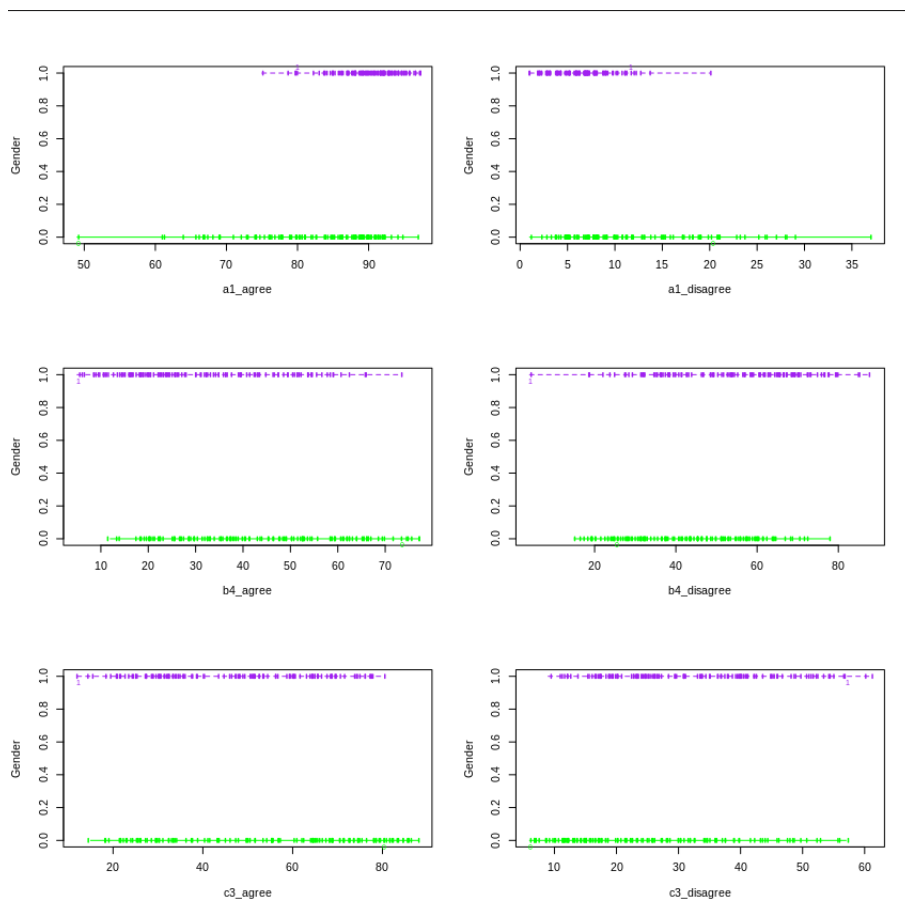
Aunque no es fácil leer una relación de la varianza y el género en la figura 4, los pesos sí señalan que rasgos correlacionados con el género, son importantes para explicar la diversidad de

los datos. Es decir, que los indicadores (las preguntas que seleccionaron para construir la encuesta) sí revelan diferencias entre género aún a escala global. Para ilustrar esto, observemos por ejemplo, las preguntas en morado sobre acuerdo o desacuerdo con frases del tipo:

- Hombres y mujeres deben de tener las mismas oportunidades en educación, trabajo, tareas domésticas, etc.
- Los hombres son los responsable de encargarse de los gastos domésticos, incluso cuando su esposa los pueda apoyar
- El rol más importante de una mujer es cuidar de la casa y de sus hijos

En estas preguntas a variables que indicaban respuestas con estereotipo se les asignaron pesos positivos y relativamente grandes, mientras que las que indicaban respuestas no estereotipadas fueron asociadas a pesos negativos y relativamente grandes. En la siguiente imagen podemos ver una distribución de las respuestas a esas preguntas, según el género:

Figura 7: Cada gráfica corresponde a una pregunta de opinión: porcentaje de respuestas afirmativas vs. género. Agrupación por género: mujeres (morado), verde (hombres).



Vemos que las mujeres tienden a: estar en mayor acuerdo con la pregunta sobre equidad; en desacuerdo con la del rol del hombre; y en una posición más o menos neutra respecto al rol de la

mujer (aunque en mayor desacuerdo en contraste con los hombres). Mientras que hay más neutralidad y desacuerdo en cuanto a la equidad en el grupo de los hombres, y posiciones más en acuerdo respecto a los roles estereotipados.

Agrupamiento por región

En la siguiente imagen se muestran los agrupamientos para las regiones de Norteamérica y Este de Asia y Pacífico. En la región de Norteamérica podemos ver claramente que los clusters separan a las observaciones por género. Por otro lado, para la región de Este de Asia no sucedía eso, así, tomamos el valor k en donde la varianza interna se reducía abruptamente (heurística del codo) para un valor de 4 grupos. En esta configuración podemos resaltar que las observaciones de hombre y mujer quedan agrupadas casi todas en el mismo cluster para cada país. Se podría pensar que tal vez el algoritmo está agrupando las observaciones dependiendo de la brecha de género que existe en cada país, es decir países con un índice de desigualdad parecido pueden encontrarse en el mismo cluster. Sin embargo, esto no es el caso debido a que según el reporte “Global Gender Gap Report 2021”¹ podemos ver que algunas observaciones que se encuentran en el mismo cluster como Australia y Japón se encuentran en posiciones muy diferentes y alejadas en el ranking.

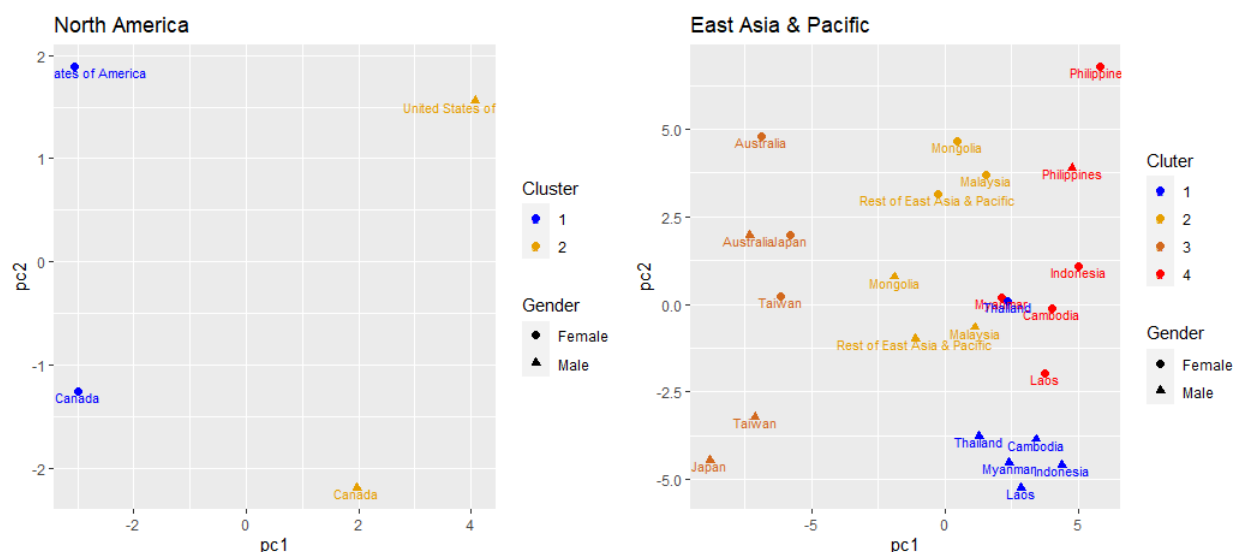


Figura 8: Agrupamiento de Regiones: Norteamérica y Este de Asia y Pacífico

Por otro lado, en el agrupamiento para la región de Latinoamérica. Podemos notar que a pesar de tener una mayor cantidad de observaciones que la región de Norteamérica, los datos quedan casi totalmente separados por género, a excepción de tres observaciones del género femenino que se encuentran en el cluster de género masculino.

¹obtenido de http://www3.weforum.org/docs/WEF_GGGR_2021.pdf

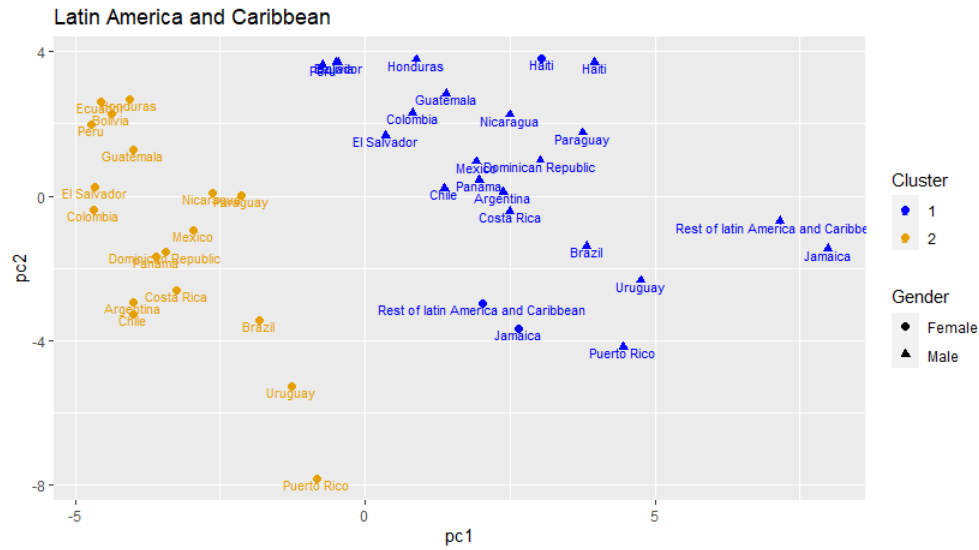


Figura 9: Agrupamiento de Región Latinoamérica y el Caribe

En las regiones del Medio Oriente y África del Norte sucede algo parecido, pero aún más marcado debido a que todas las observaciones quedan agrupadas completamente por género. Esto es algo que no resulta tan impactante al considerar la desigualdad que se sabe que se vive en ciertos países de la región. Podemos incluso resaltar que observaciones como Israel se encuentran más cerca en la proyección, lo cual concuerda con la menor brecha de género que existe en este país si miramos el ranking mencionado anteriormente del reporte “Global Gender Gap Report 2021”.

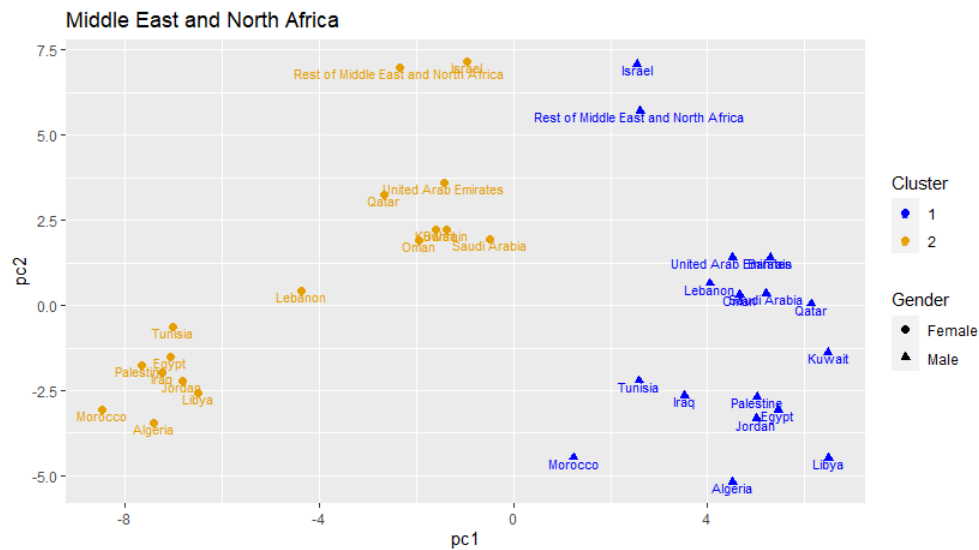


Figura 10: Agrupamiento de Región Medio Oriente y África del Norte

La región de Europa y Asia Central fue probablemente la región en donde más observaciones de género masculino y femenino se mezclaban. Esto resulta coherente e incluso esperado tomando en

cuenta que muchos países de Europa se encuentran en las mejores posiciones del ranking de reducción de la brecha de género. En la siguiente imagen en donde se muestran 4 grupos podemos resaltar que la mayoría de las observaciones del cluster anaranjado engloban tanto el género masculino como el género femenino de varios países y se encuentran entre las primeras posiciones del ranking mencionado anteriormente. Sin embargo, es importante mencionar que esto no sucede para todos los países como Francia y España, países que se encuentran en las mejores 16 posiciones del ranking. Esto podría deberse a un desbalance en la cantidad de personas de ambos géneros que llenaron la encuesta. También es importante considerar que la encuesta tiene muchas preguntas que son de índole muy subjetiva y esto puede sesgar los resultados. O simplemente puede deberse a que esta cantidad de clusters no refleja en realidad la naturaleza de los datos.

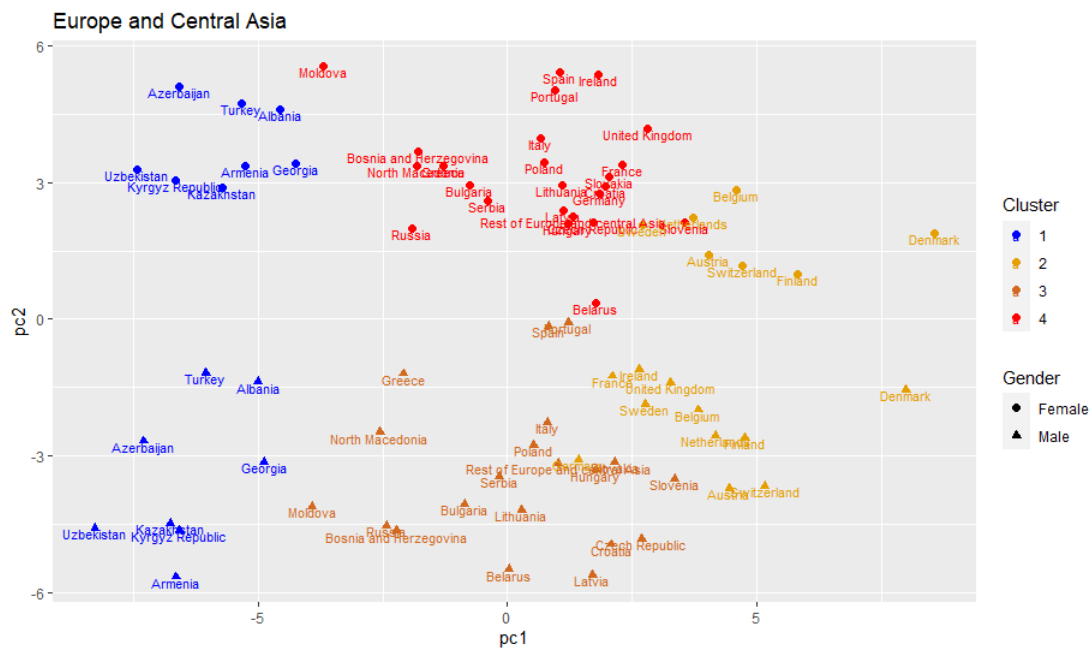


Figura 11: Agrupamiento de Región de Europa y Asia Central

Por último podemos ver en la siguiente imagen que con la región de Sub-Saharan África sucede algo similar que con la región de Latinoamérica, que la mayoría de las observaciones se encuentran separadas por género y únicamente 3 observaciones que se encuentran en la frontera entre ambos clusters se encuentran con el género opuesto.

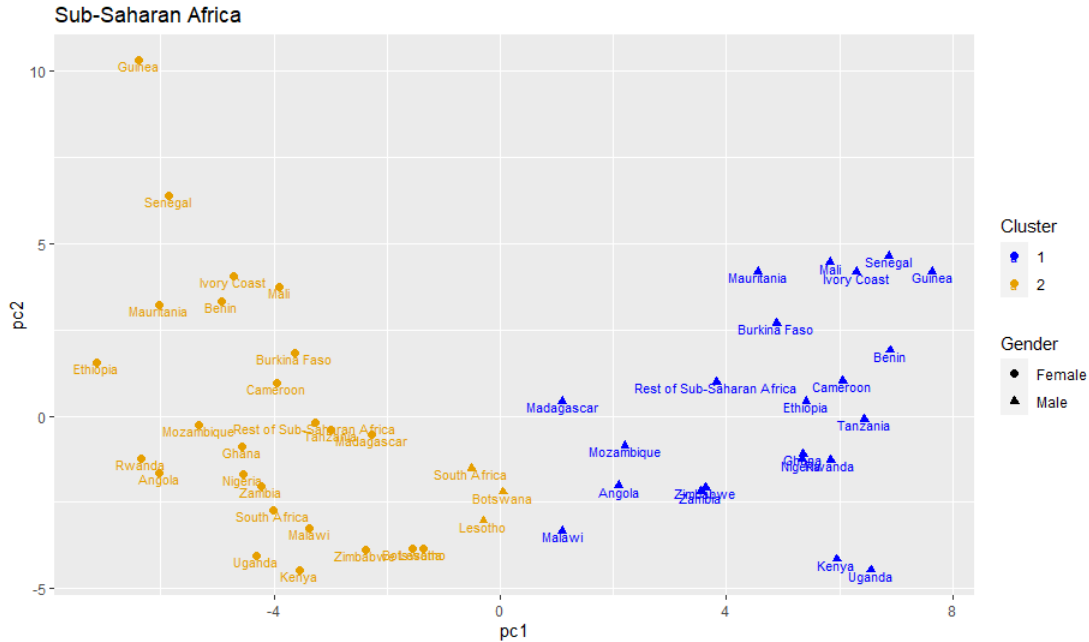


Figura 12: Agrupamiento de Región de Sub-Saharan África

A través de las gráficas anteriores, notamos que en diversas regiones la desigualdad de género es algo que parece estar presente debido a como los datos fueron agrupados en estos casos por el género de las observaciones. Al ver los datos podemos notar que el género en cuestión es una de las variables para cada observación, pero cabe resaltar que esta variable fue ignorada al momento de realizar la proyección de los datos a través de PCA y también fue ignorada para realizar el agrupamiento de los datos. Por lo tanto, se tiene una fuerte sospecha para asegurar que las gráficas y proyecciones en efecto evidencian que la desigualdad de género es un factor relevante en la naturaleza de los datos.

III. Relaciones entre variables y desbalance en trabajo

Modelos de Predicción

Las técnicas de predicción pueden ser utilizadas con diversos fines, uno de ellos muy común es predecir a través de una regresión o una clasificación un valor de interés para tomar de manera activa una acción con base en esta predicción. Un ejemplo podría ser predecir el tipo de cliente que puede ser una persona para una empresa para así poder dirigir publicidad específica a ese cliente. Sin embargo, en este caso la predicción no tiene este fin, ya que no tendría mucho sentido predecir el género de las observaciones para poder tomar una acción directa ya que de por sí las observaciones son un promedio de las respuestas de las personas de cada país. Mas bien, el fin de clasificar el género correctamente en este escenario es entender la naturaleza de estos datos y comprender que predictores pueden ser relevantes para evaluar la desigualdad de género existente.

Por ello, cómo algoritmos de clasificación se eligieron el de Regresión logística y el de Árboles de decisión. Ya que ambos son interpretables y podrían mostrar relaciones entre las múltiples variables.

Regresión Logística y LDA

La regresión logística es una de las técnicas de predicción más utilizadas para los sistemas de clasificación binaria, por lo tanto la podemos utilizar en este contexto ya que lo que nos interesa predecir es el género de las observaciones (1 si es mujer, 0 si es hombre). LDA por otro lado es un clasificador que es muy efectivo cuando los datos son linealmente separables. A continuación detallaremos diferentes configuraciones utilizadas para evaluar la naturaleza de los datos y ver que predictores pueden revelar información importante.

El primer modelo que se intentó ajustar fue un modelo en donde se utilizó una combinación lineal de todos los predictores. Sin embargo para este modelo no se logró conseguir la convergencia. Esto podría sugerir que los datos viven en un espacio en donde son linealmente separables, lo cual evidenciaría fuertemente que existe una brecha de género muy marcada en los datos.

Esto nos motivó entonces a ajustar un modelo LDA para todos los predictores. Dividimos el conjunto de datos en 80 % como conjunto de entrenamiento y 20 % como conjunto de prueba y ejecutamos el experimento 100 veces. Obtuvimos una precisión promedio del 98 %. De esta forma podemos ver que hay evidencia para considerar que los datos son linealmente separables. Esto incluso se puede ver reforzado si consideramos las gráficas que vimos en la sección de agrupamiento dado que en varias regiones se podía ver que los datos quedaban separados por género y en efecto eran linealmente separables.

Como planteamos inicialmente, es de interés comprender que factores pueden ser de mayor relevancia para detectar la desigualdad de género en un país. Para ello ajustamos diversos modelos de regresión logística considerando solo cierto grupo de predictores para evaluar el desempeño de cada clasificador y compararlo con el modelo anterior. Entre mejor sea el desempeño, tendremos una mayor razón para considerar que existe una fuerte relación entre esos predictores y la desigualdad de género existente.

El mejor modelo que se logró ajustar fue un modelo que consideraba únicamente predictores relacionados con el trabajo en casa. Utilizamos los siguientes predictores: 'c2_animals', 'c2_clean', 'c2_cook', 'c2_shop' y 'c2_hours_chores'. Con estos predictores obtuvimos un modelo que se ajusta al modelo saturado bastante bien, pues obtenemos un valor de residual deviance de 27.5 con 240 grados de libertad, lo cual indica que tenemos un buen modelo suponiendo que el modelo saturado es un buen ajuste. En este modelo no hay evidencia para rechazar alguno de estos predictores utilizados. Luego ejecutamos 100 experimentos con este modelo en donde se dividían el conjunto de datos en 80 % como conjunto de entrenamiento y 20 % como conjunto de prueba. De estas 100 ejecuciones obtuvimos un promedio de 97 % de precisión. Podemos ver entonces que estos predictores en efecto revelan información importante para predecir el género de una observación, y por lo tanto hay razones para considerar que estas variables marcan una gran diferencia en la brecha de género.

Otro modelo que se ajustó bastante bien fue un modelo basado en predictores relacionados con el aspecto económico y financiero. Para este modelo se utilizaron los predictores: 'b2_self', 'b6_wage', 'b6_non_wage', 'b3_fullyindependent', 'b5_man_other'. Obtuvimos un valor de residual deviance de 118 con 240 grados de libertad. Podemos ver que esto refleja un buen ajuste respecto al modelo

saturado. Utilizando el mismo acercamiento descrito anteriormente obtuvimos una precisión de predicción del 90 % sobre el conjunto de prueba.

Se probó ajustar un modelo también con variables relacionadas a la opinión y estereotipos culturales, y el ajuste obtenido tiene una menor bondad de ajuste respecto a los dos modelos anteriores. Esto parece indicar que las opiniones y estereotipos culturales no son tan importantes para estudiar la naturaleza de los datos en comparación con las variables de tipo financieras y trabajo doméstico.

Por último se probó ajustar también un modelo con variables que pueden ser consideradas como "desventajas" para los hombres como trabajos de mayor carga física o mayor carga laboral para ver si eso explicaba la poca participación de los hombres en labores domésticas, sin embargo el modelo ajustado no reveló una bondad de ajuste tan alta como los primeros dos modelos, y es incluso menor que el ajuste obtenido a través de la opinión y estereotipos culturales. Esto nuevamente sugiere que los factores más relevantes para analizar la brecha de género se encuentra en variables que son desventajas para las mujeres como los trabajos domésticos y aspectos financieros y de independencia económica.

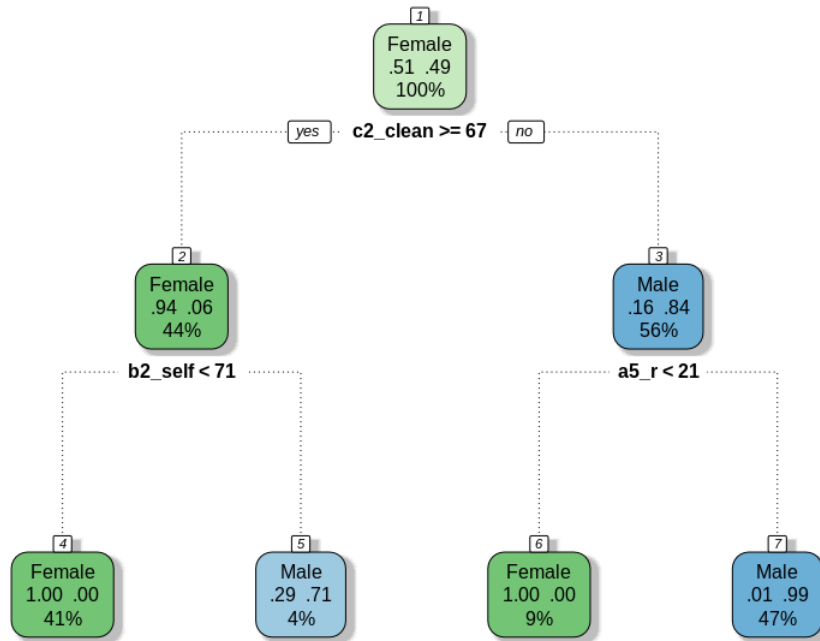
Tabla 1: Resultados promedios de validación cruzada con $k = 15$ para distintos predictores.

| Classifier | macro Precision | macro Recall | macro F1 |
|-------------------|------------------------|---------------------|-----------------|
| RL House-chores F | 0.97 | 0.96 | 0.96 |
| RL Economics | 0.89 | 0.89 | 0.89 |

Árboles de decisión

Para elegir un árbol óptimo se dividió de nuevo el conjunto en 80 % para entrenamiento y el resto para test. En la figura 13 se observa el árbol óptimo; para construirlo se contruyó el árbol completo (con el parámetro $\alpha = 0$) el $\alpha = 0.031$ óptimo se obtuvo por validación cruzada (figura 14).

Figura 13: Árbol óptimo de decisión

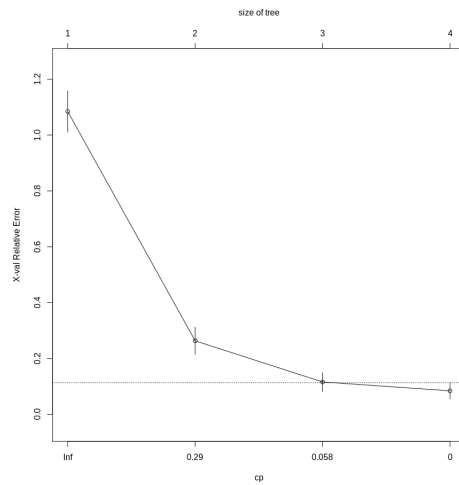


En el árbol se puede observar que la variable **c2_clean**, es uno de los nodos que más fuertemente reduce la variabilidad interna de las regiones. Específicamente, los nodos corresponden a las preguntas que se enlistan:

1. De manera habitual, ¿te haces cargo de la limpieza en tu casa al menos la mitad de las veces?
2. ¿Durante el último año has sido predominantemente totalmente independiente financieramente?
3. ¿Eres el/la principal responsable de tomar las decisiones y aportar dinero a tu hogar?

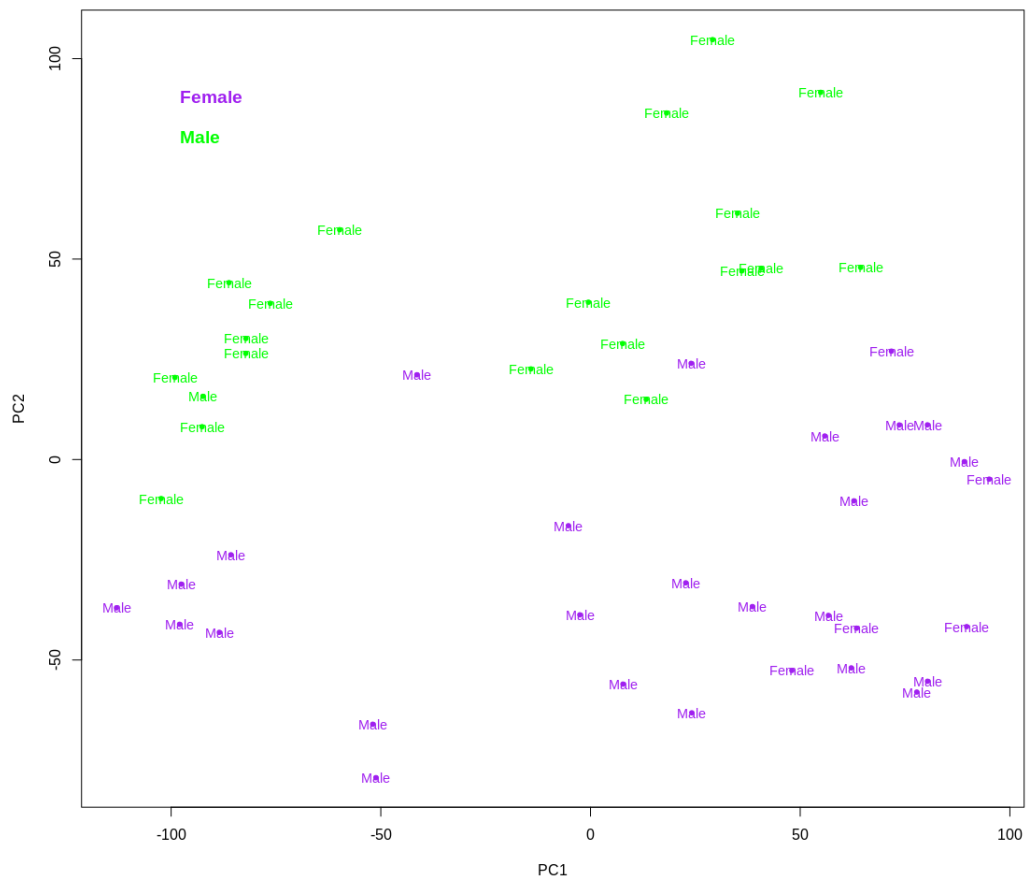
Si observamos los nodos terminales, podemos notar que en los que contienen la mayoría de los datos (el 41 % y el 47 %) la clasificación es casi perfecta. Corresponde uno casi a únicamente mujeres (verde) y el otro a hombres (azul). En el conjunto de prueba la precisión fue del 96 %. Siguiendo el flujo de preguntas, notamos que las mujeres suelen limpiar más que los hombres, los hombres tienden a ser más independientes financieramente y son los los responsables de tomar decisiones y aportar dinero en sus casas.

Figura 14: Validación cruzada para decidir el parámetro α .



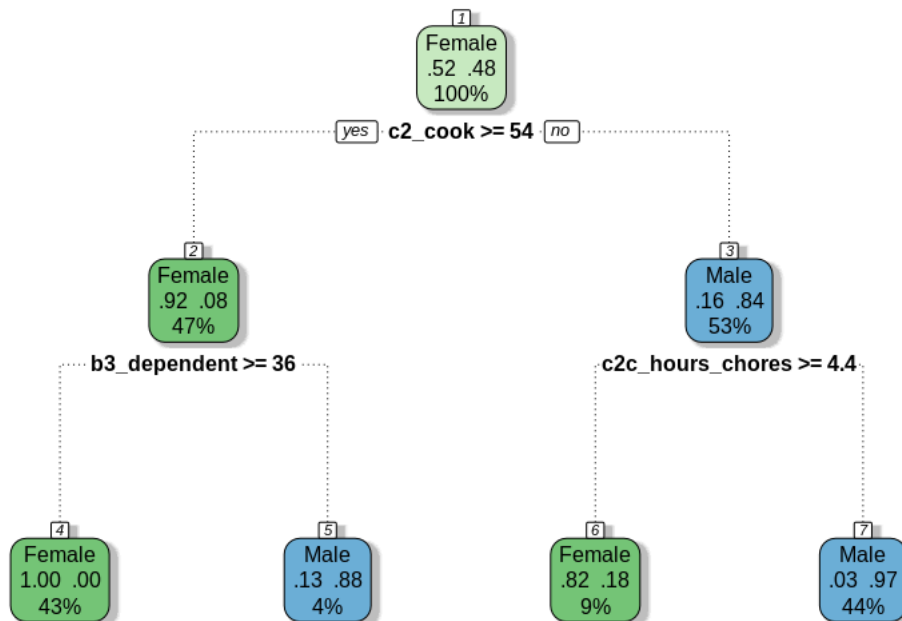
También se muestra una visualización de la clasificación, utilizando la proyección en componentes principales, de los países-género.

Figura 15: Clasificación de árbol óptimo.



Ahora, omitiendo los nodos anteriores y entrenando otro árbol, podemos encontrar otras preguntas útiles para distinguir entre los dos géneros. Esto es interesante ya que como varias preguntas pueden estar correlacionadas es posible obtener otros clasificadores con precisiones similares. Además, los árboles son sensibles a cambios en las muestras. Volviendo a tomar otro conjunto de entrenamiento del mismo tamaño, y quitando los nodos anteriores construimos otro árbol de decisión.

Figura 16: Árbol óptimo de decisión



Efecitivamente, los nodos elegidos están naturalmente correlacionados; la pregunta sobre limpieza cambió por una pregunta análoga pero relacionada con la cocina, la pregunta sobre independencia económica cambió por una pregunta sobre dependencia económica, además de que el nodo 3, fue reemplazado por un promedio del número de horas que se dedican a tareas del hogar. De nuevo, siguiendo la proporción de los géneros en los nodos terminales, observamos que un criterio para clasificar con precisión entre hombres y mujeres (del 87 % en este caso) tiene que ver con tareas del hogar, donde la mujeres participan mayormente y con cuestiones financieras, las mujeres presentan más dependencia económica.

Tabla 2: Resultados promedios de validación cruzada con $k = 15$ para los dos árboles.

| Classifier | Precision | Accuracy | Recall | macro F1 |
|-----------------|-----------|----------|--------|----------|
| Decision Tree 1 | 0.97 | 0.96 | 0.96 | 0.96 |
| Decision Tree 2 | 0.97 | 0.96 | 0.96 | 0.96 |

Conclusiones

- Podemos concluir, que al menos en la población con acceso a internet, hay un desbalance entre el tiempo que dedican mujeres y hombres a tareas domésticas en escala global. Donde los que gozan de menores responsabilidades y tiempo invertido en tareas domésticas son los hombres. Estas diferencias en el tiempo trabajado en tareas domésticas no se explican a través de una mayor incidencia de los hombres en otro tipo de trabajos, por lo cual se tiene la teoría que la desigualdad en esta área se debe mayormente a factores culturales. Los algoritmos de agrupamiento parecen indicar que aunque se observan diferencias en la mayoría de las partes del mundo respecto al género según los predictores estudiados, esa distinción no es homogénea en distintas regiones.
- Existe una asociación con opiniones frente al rol de género. Un punto interesante es que las mujeres tienden a estar de acuerdo frente a la equidad de género y simultáneamente parecen tener una opinión neutra en cuanto el rol estereotipado de la mujer frente a tareas del hogar (es decir no se observa una tendencia a estar en desacuerdo). Además, suelen estar en desacuerdo frente al rol del hombre como protector económico. Por otro lado, los hombres tienden a estar de acuerdo con los roles tradicionales de hombres y mujeres impuestos por la cultura.
- Parece haber una correlación positiva entre dedicar menos tiempo al hogar y realizar menos trabajo remunerado. Asimismo, también parece haber una correlación positiva entre dedicar más tiempo al hogar y ser menos independiente económicamente. Esto resulta natural y coherente si consideramos los tiempos requeridos para la realización de tareas domésticas.
- Es interesante notar que para el algoritmo de árbol bastó realizar 3 preguntas para tener clasificaciones con accuracy alrededor del 90 % en clasificación de géneros. Lo cual es indicativo de que tareas del hogar e independencia económica es algo que marca fuertemente la brecha de género. Esta misma conclusión se puede observar también si analizamos los modelos ajustados basados en regresiones logísticas en donde los modelos con una mejor bondad de ajuste eran justamente los modelos relacionados a predictores del trabajo doméstico y de la independencia económica y sector financiero.