

**UNIVERSIDAD DE CHILE
FACULTAD DE MEDICINA
ESCUELA DE SALUD PÚBLICA**



**APLICACIÓN DE METODOS MULTIVARIADOS PARA EL
ESTUDIO DE LA ASOCIACION DE FACTORES DE
RIESGO CARDIOVASCULAR Y DETERMINANTES
SOCIALES: EVOLUCION EN EL TIEMPO EN CHILE**

CLAUDIA CHAVEZ OYANADEL

**TESIS PARA OPTAR AL GRADO DE
MAGISTER EN BIOESTADISTICA**

**Director de Tesis: Dr. Sergio Muñoz N.
Profesor Co-Patrocinador: Dr. Shrikant Bangdiwala**

Santiago, Diciembre 2015

**INFORME DE APROBACION TESIS DE
MAGISTER EN BIOESTADISTICA**

Se informa a la Comisión de Grados Académicos de la Facultad de Medicina, que la Tesis de Magíster en Bioestadística presentada por la candidata

CLAUDIA CHAVEZ OYANADEL

ha sido aprobada por la Comisión Informante de Tesis como requisito para optar al Grado de **MAGISTER EN BIOESTADISTICA** en Examen de Defensa de Tesis rendido el día **01 de Diciembre del 2015**

Dr. Sergio Muñoz N.
Director de Tesis

Escuela de Salud Pública, Facultad de Medicina, Universidad de Chile

Dr. Shrikant Bangdiwala
Profesor Co-Patrocinador

COMISION INFORMANTE DE TESIS

Dr. Sergio Alvarado O.

Sr. Gabriel Cavada Ch.

Prof. Dr. Mauricio Canals L.
Presidente Comisión de Examen

Índice

1	Capítulo I: Introducción.....	1
2	Capítulo II: Marco Teórico de Referencia	3
3	Capítulo III: Objetivos Generales y Específicos.....	16
4	Capítulo IV: Hipótesis.....	18
5	Capítulo V: Metodología.....	19
5.1	Metodología Estadística	19
5.1.1	BIPLOT Clásicos.....	19
5.1.2	HJ BIPLOT.....	23
5.1.3	BIPLOT Componentes Principales.....	24
5.1.4	BIPLOT Análisis de Correspondencias.....	25
5.1.5	BIPLOT de Regresión.....	26
5.1.6	Características de los marcadores.....	27
5.1.7	Calidad de la representación	30
5.1.8	Variables suplementarias en BIPLOT.....	31
5.2	Metodología del estudio	32
5.2.1	Universo y Muestra.....	32
5.2.2	Variables o tópicos de estudio. Fuentes de información	34
5.2.3	Métodos de recolección de la información e instrumentos usados	42
5.2.4	Técnicas de procesamiento de datos	43
5.2.5	Análisis e interpretación	46
5.2.6	Limitaciones.....	48
5.2.7	Aspectos éticos	49
6	Capítulo VI: Resultados.....	50
6.1	Análisis descriptivos y reducción de variables.....	50
6.2	Asociación factores riesgo ECV y determinantes sociales.....	60
7	Capítulo VII: Discusión o Comentarios.....	80
8	Capítulo VIII: Conclusiones	83
9	Capítulo IX: Bibliografía	84

Dedicatoria

A mi esposo Cristián Baeza

A mi bebe Matías

A mi madre, padre y hermanos

Agradecimientos

Agradezco a la Escuela de Salud Pública de la Universidad de Chile que me brindó la oportunidad de enriquecer mi carrera profesional e iniciar una carrera en la investigación. En especial al Dr. Sergio Muñoz y Dr. Shrikant Bangdiwala por creer en mí, confiar plenamente en mi trabajo, y brindarme apoyo incondicional.

Agradezco además la eficiencia y respuesta oportuna del profesor José Luis Vicente Villardón de la Universidad de Salamanca, quién actualizó su software incorporando parte fundamental incluida en esta tesis.

Finalmente, un agradecimiento muy especial a mi esposo Cristián Baeza, quien me acompañó en este recorrido y me entendió en todo momento, incluso cuando además de estar estudiando estábamos esperando a nuestro pequeño Matías. A mi madre, por una vez más, estar conmigo en las instancias difíciles apoyándome totalmente. Y a mi familia completa que me aportan cada uno con un granito de arena.

Resumen

En Chile, el infarto al miocardio es la primera causa de muerte específica y la enfermedad cerebrovascular es una de las principales causas de discapacidad y deterioro de la calidad de vida. En esta investigación, se aplican técnicas de análisis multivariado (BILOT) que permitan explicar la asociación entre un conjunto de factores de riesgo cardiovascular y determinantes sociales basados en la información recogida periódicamente en Chile a nivel poblacional mediante encuestas de salud (ENS) y de caracterización socioeconómica (CASEN) en los años 2009 y 2003, y analizar si la estructura de las asociaciones encontradas tiene variación temporal. Los métodos BILOT buscan resumir en una sola imagen las correlaciones entre las variables de estudio, la variabilidad de los mismos y la similitud de las observaciones, en este caso las regiones del país. Los resultados obtenidos son consistentes con otras investigaciones y los métodos BILOT describen la asociación entre factores de RCV y determinantes sociales de manera óptima y resumida. Además, la incorporación de variables suplementarias complementa la interpretación gráfica y mejora la toma de decisiones.

Palabras claves: BILOT, RCV, Det. Sociales, Análisis Componentes Principales, Variables suplementarias.

1 Capítulo I: Introducción

Las enfermedades cardiovasculares (ECV) son la causa de muerte y discapacidad más importantes en el mundo, Latinoamérica y en Chile (1,2).

Los factores de riesgo de ECV son conocidos y su relación con determinantes sociales también, esta asociación ha sido estudiada mediante análisis estadísticos clásicos en Bioestadística, tales como análisis de sobrevivencia, análisis de regresión, y métodos de análisis descriptivo exploratorio, mediante tablas de contingencia y análisis estratificado. Todos estos métodos consideran una variable de respuesta que puede estar definida como un factor de riesgo cardiovascular o como una medida de resumen del conjunto de factores estudiados.

En el área de la Salud Pública el uso de análisis multivariado, así como sus representaciones gráficas mediante BIPLLOT, no han sido suficientemente explorados en comparación a otras disciplinas como, Ecología o Genética (3-7). La potencialidad de este tipo de representaciones gráficas radica en el hecho de poder visualizar relaciones no solo entre conjuntos de variables sino también entre los sujetos de estudio a los que se realiza el análisis, simultáneamente (8).

Esta investigación propone estudiar y representar la asociación entre dos conjuntos de variables: el conjunto de factores de riesgo cardiovascular y el conjunto de determinantes sociales, mediante la información recogida en encuestas

poblacionales realizadas en Chile a nivel regional mediante metodologías de análisis multivariado y representación gráfica utilizando BIPLLOT.

Con el uso de esta metodología se espera representar simultáneamente la asociación entre estos dos conjuntos de variables, además de identificar descriptivamente cambios de dicha asociación entre dos momentos en el tiempo a nivel regional y de manera descriptiva. Se espera que ésta gráfica sea lo suficientemente clara como para que profesionales alejados de la estadística logren identificar y comprender los resultados representados.

2 Capítulo II: Marco Teórico de Referencia

Enfermedades cardiovasculares y determinantes sociales en Chile e instrumentos de medición poblacional

Las enfermedades cardiovasculares (ECV) son la primera causa de muerte y discapacidad a nivel mundial, en efecto corresponden al 30% del total de las muertes registradas en el mundo (9), cifra que se proyecta se mantendrá inalterada en los próximos años (10).

Si bien Chile es uno de los países que tiene menores índices de ECV (1), el infarto al miocardio es la primera causa de muerte específica, y la enfermedad cerebrovascular es una de las principales causas de discapacidad y deterioro de la calidad de vida (2), con esto las ECV son importantes problemas de Salud Pública. Numerosos estudios y reportes muestran que los factores de riesgo más relevantes de ECV son: diabetes, hipertensión, colesterol elevado, consumo de tabaco, mala alimentación, e inactividad física, entre otros (11-15). Por otro lado, se sabe que una serie de determinantes sociales modifican indirectamente el riesgo de ECV, dentro de los cuales se destaca, la distribución de ingreso y el nivel educacional, además de las condiciones de vida, índices de alfabetización, y el sistema de salud (16-20). De aquí la importancia de identificar las asociaciones entre los determinantes sociales y los factores de riesgo de ECV y representarlas simultáneamente con el fin de obtener la mayor cantidad de información posible para la evaluación y desarrollo de políticas públicas, y el entendimiento del problema.

Con el fin de complementar las conclusiones en esta investigación, y en vías a la generación de políticas públicas preventivas se agregan dos variables asociadas al riesgo cardiovascular; la prehipertensión y la prediabetes, en busca de identificar las regiones que presentan mayores prevalencias y con qué determinantes sociales y/o riesgo cardiovascular se asocian.

El concepto de prehipertensión fue incorporado en el año 2003 mediante el VII reporte del Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure de los Estados Unidos de Norteamérica, en el que se define el diagnóstico de prehipertensión para aquellos pacientes que reiteradamente presentan niveles de presión arterial sistólica entre 120-139 mg/dl y/o presión arterial diastólica entre 80-89 mg/dl (21).

Por su parte, en el documento “Consenso de Prediabetes” publicado por la Asociación Latinoamericana de Diabetes (ALAD) se llegó a acuerdo de que niveles de glucosa plasmática entre 140 mg/dl y 199 mg/dl (Tolerancia a la glucosa alterada (TGA)) medido 2 horas después de una carga de glucosa, y niveles de glucosa plasmática luego de un ayuno de 8 horas (Glucosa Alterada en Ayuno (GAA)), se encuentran entre 100 mg/dl y 125 mg/dl se consideran valores razonables para diagnosticar prediabetes (22).

Considerando que tanto la prediabetes como la prehipertensión son la antesala para padecer problemas de salud que se consideran como factor de riesgo cardiovascular es que se justifica por sí solo la incorporación de estas variables.

En Chile, oficialmente el Ministerio de Salud, realiza encuestas poblacionales cuyo objetivo es diagnosticar la salud de la población chilena y con esto orientar las políticas de salud. Estas encuestas se realizan periódicamente, a saber¹:

- Año 2000, primera Encuesta Nacional de Calidad de Vida y Salud (ENCAVI).
- Año 2003, primera Encuesta Nacional de Salud (ENS)
- Año 2006, segunda ENCAVI, y
- Año 2009, la segunda ENS.

ENCAVI, es la encuesta previa a la ENS y abarca preguntas referentes a aspectos sociales y de salud, esta encuesta reporta información acerca de ciertas patologías diagnosticadas (mediante auto reporte), e información referente a alcoholismo, tabaquismo, sedentarismo, y alimentación, entre otros (23).

La ENS mide patologías y/o condiciones de salud diversos, incluyendo los factores de riesgo de enfermedades cardiovasculares descritos en párrafos anteriores, tales como, hipertensión arterial, dislipidemias, tabaquismo, obesidad, diabetes mellitus, sedentarismo, sintomatología y riesgo cardiovascular, estado nutricional, entre otros. Estos indicadores se midieron en ambas aplicaciones, sea la del año 2003 como la del año 2009 (24).

Por su parte, el Ministerio de Desarrollo Social (ex MIDEPLAN) desarrolla la Encuesta de Caracterización Socioeconómica Nacional (CASEN), la que tiene por

¹ <http://epi.minsal.cl/estudios-y-encuestas-poblacionales/encuestas-poblacionales/>

objetivo conocer periódicamente la situación de los hogares y de la población en relación a aspectos demográficos, de educación, salud, vivienda, trabajo e ingresos, y evaluar el impacto de la política social en el país. Además, esta encuesta busca estimar la cobertura, focalización y distribución del gasto fiscal utilizado en los programas sociales más importantes para el país. Esta encuesta se ha realizado periódicamente a partir del año 1985 con una periodicidad bianual o trianual, teniendo una versión para 2003 y 2009 que interesa en este estudio².

La encuesta CASEN consta de módulos referentes a registro de residentes, trabajo e ingresos, en donde se realizan preguntas pertinentes a la construcción de indicadores sociales.

Utilizando los indicadores desarrollados a nivel regional, por ENS y CASEN, se aplicarán metodologías de análisis multivariado y representaciones gráficas BIPLLOT que permitan representar simultáneamente la asociación existente entre un conjunto de factores de riesgo cardiovascular (ENS) con un conjunto de determinantes sociales (CASEN), y su evolución en el tiempo para las regiones de Chile.

² http://observatorio.ministeriodesarrollosocial.gob.cl/casen_obj.php

Métodos multivariados BIPLLOT

Los gráficos BIPLLOT comienzan con los estudios de Gabriel (25), donde se define que el objetivo principal de estos análisis es describir una matriz rectangular en una representación gráfica de baja dimensión. Esta representación no solo muestra la información resumida, sino que también las correlaciones existentes entre individuos y variabilidad de manera simultánea (4). Las representaciones gráficas BIPLLOT no solo se utilizan para describir una matriz de datos, sino que también de manera posterior a un análisis multivariado ad-hoc a las matrices de datos con las que se está trabajando (26).

La importancia de este tipo de gráficos radica en que su interpretación se basa en conceptos geométricos sencillos, tales como; la similitud entre los individuos, que es una función inversa de las distancias de los mismos; las longitudes y ángulos de los vectores representados en un BIPLLOT que se encuentran en función de la variabilidad y covariabilidad de las variables; y la relación entre individuos y variables que se interpretan según las proyecciones de los puntos (individuos) sobre los vectores (variables) (27).

La forma más sencilla de un BIPLLOT es un gráfico de dispersión, en el cual las variables son representadas por los ejes y las observaciones por los puntos en el plano (27). Sin embargo, cuando es necesario representar más de dos variables esta representación se complica y se requiere transformaciones de los datos para lograr su representación.

Los BIPLLOT no necesariamente ofrecen un análisis de variables latentes, como es el caso de componentes principales, sino que corresponde a un enfoque complementario de representación de matrices de datos (28).

De manera general, en las gráficas BIPLLOT se interpretan las distancias entre observaciones o unidades muestrales como disimilaridades entre los mismos, las longitudes de los vectores representan la variabilidad de la variable, el ángulo que forman los vectores de variables se interpreta en términos de correlación y la relación observación-variable se estudia a través de la proyección de los puntos sobre los vectores que representan las variables (27).

Por ejemplo, en el trabajo de Díaz-Faes (29) busca representar centros de investigación y su actividad científica mediante análisis BIPLLOT, en su documento expone una gráfica que explica claramente cómo interpretar este tipo de gráficos. La figura 1 muestra claramente que los centros corresponden a los puntos en el gráfico y los vectores a las variables. Así las distancias entre los puntos son similitudes entre los centros (centro 1 similar a centro 3), los vectores corresponden a las variables bibliométricas; por lo que se ve que la variable 3 es la que presenta mayor variabilidad y las variables 1 y 2 se correlacionan de manera directa, además de que el centro 1 presenta valores más altos de la variable 1 que el centro 2.

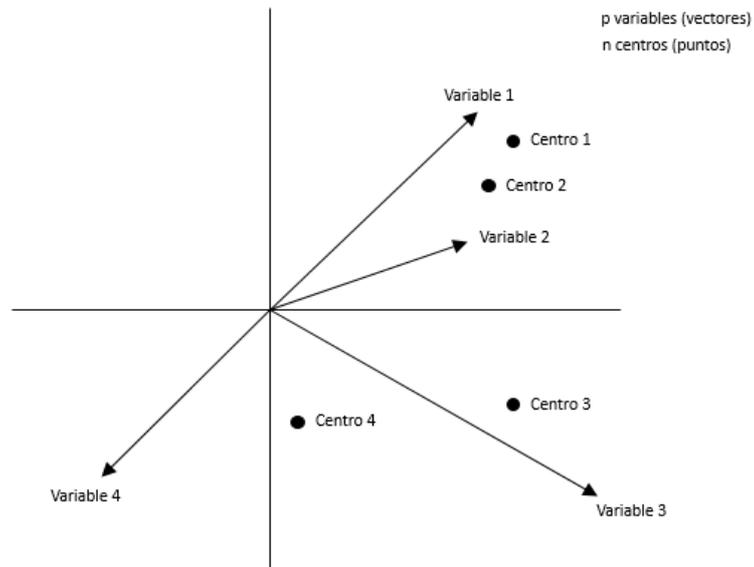


Figura 1. Interpretación Biplot (Díaz-Faes, 2013)

De manera muy general, el método BIPLLOT busca aproximar una matriz de datos X de n observaciones y p variables ($n > p$), mediante una descomposición de valores singulares (DVS), tal como (4, 8, 27, 28):

$$X \cong UDV' \quad (2.1)$$

Donde U y V son matrices singulares ortogonales tales que $U'U = V'V = I$ (I identidad), y D corresponde a la diagonal que contiene los α_k mayores valores singulares. Esta descomposición se puede escribir como:

$$X \cong (UD^s)(D^{1-s}V') = AB' \quad (2.2)$$

Donde A y B son las matrices que contienen a los marcadores fila y columna que se utilizarán simultáneamente en el gráfico (4, 8, 27, 28).

Existen dos tipos principales de BIPLLOT (28):

- Los BIPLLOT simétricos, que entregan información de filas y columnas de tablas de doble entrada o de una matriz de datos. En este tipo de BIPLLOT la información puede ser intercambiada sin pérdida alguna de información esto es intercambiar filas por columnas en la matriz de datos o la tabla de contingencia.
- Los BIPLLOT asimétricos, que entregan información sobre las unidades de la muestra o las variables de la matriz de datos. Dado que las variables y las unidades muestrales son muy distintas en su representación, los objetos del gráfico no pueden ser intercambiados.

Por otro lado, los gráficos BIPLLOT se pueden clasificar en una gran variedad dependiendo de sus propiedades, siendo los más importantes (4); BIPLLOT clásicos, se encuentran GH BIPLLOT, JK BIPLLOT, SQRT BIPLLOT, propuestos por Gabriel, y HJ BIPLLOT propuesto por Galindo; BIPLLOT de regresión; BIPLLOT basados en componentes principales; BIPLLOT basados en análisis de correspondencias, entre otros³.

Los BIPLLOT clásicos son aquellos que aproximan una matriz sin hacer supuestos sobre las distribuciones probabilísticas y se utilizan principalmente de manera de

³ El detalle estadístico de cada uno de ellos se presenta en el capítulo 5, específicamente en la sección 5.1. Metodología Estadística.

describir una matriz de datos (30). Además, dependiendo de su tipo de descomposición o selección de marcadores, se clasifican en:

- **GH BILOT**, $s = 0$ en (2.2): Este tipo de BILOT se basa en los marcadores columna, basándose en la restricción $U'U = I$. Por lo tanto, este tipo de BILOT consigue una alta calidad de variables y una baja de los individuos (29). Este tipo de BILOT es denominado CMP-BILOT (Column Metric Preserving) por Greenacre en el año 1984 (8,31).
- **JK BILOT**, $s = 1$ en (2.2): Este tipo de BILOT se centra en los marcadores fila, basándose en la restricción $V'V = I$. Esta vez la mejor representación es de los individuos y no de las variables (30). De la misma forma anterior, Greenacre en 1984 lo denomina RMP-BILOT (Row Metric Preserving) (8,31).

Ninguna de las dos representaciones anteriores logra obtener la misma bondad de representación para marcadores fila como para marcadores columna (30). Para mejorar este problema, dentro de los BILOT clásicos se encuentra SQRT BILOT y una extensión denominada HJ BILOT.

- **SQRT BILOT**, $s = 1/2$ en (2.2): Este tipo de BILOT representa tanto filas y columnas, ya que su factorización no es única ($A'A \neq I$ y $B'B \neq I$). Este tipo de análisis es utilizado principalmente para la aproximación de una matriz X , como es el caso de la diagnosis de modelos en tablas de contingencia (4).

Durante el año 1986 Galindo propone una extensión de los BIPLLOT clásicos, en los cuales sea posible representar los marcadores fila y columna con la máxima calidad posible (31). Este tipo de BIPLLOT se denomina HJ BIPLLOT o RCMP-BIPLLOT. Este tipo de BIPLLOT se utiliza en la interpretación simultánea de variables y observaciones, y su interpretación es muy similar al análisis de correspondencia, pero incorporando variables numéricas (28,30).

- **HJ BIPLLOT** seleccionan la factorización $X \cong (UD)(DV')$ para su representación consiguiendo una alta calidad. La desventaja que tiene este método es que no representa directamente la matriz X (4). Sin embargo, su gran ventaja y mayor utilización se centra en que posee las mismas ventajas del análisis de correspondencia, pero es aplicable a cualquier matriz de datos y no solo a las referidas a frecuencias (29).

La mayoría de los métodos BIPLLOT revisados en los párrafos anteriores, utilizan distancias pitagóricas para la representación, en busca de obtener una representación más general, Vicente-Villardón define los BIPLLOT Generalizados, lo que permite incluir métricas a los ya elaborados por Gabriel y Greenacre (30).

Además de los BIPLLOT clásicos, se encuentran aquellos que se desprenden de análisis multivariados, cuyo enfoque es una particularidad de los BIPLLOT Generalizados; como los basados en componentes principales, análisis de correspondencia, y regresión, entre otros.

- **BIPLOT basados en componentes principales (ACP):** Los BIPLLOT basados en componentes principales son considerados los BIPLLOT asimétricos más simples y populares (8). Este tipo de análisis se realiza en caso de contar con variables continuas y su finalidad radica en el análisis de las relaciones entre individuos y/o variables a través de componentes principales o variables latentes (32).

Representan la información centrada mediante una aproximación de valores singulares (DVS) y en ocasiones variables estandarizadas, lo que entrega una ventaja en la interpretación de las gráficas obtenidas para los factores encontrados (8, 28). Si dicha descomposición es por filas o columnas, similar a lo visto de BIPLLOT clásicos, se encuentran dos tipos de BIPLLOT diferentes; RMP y CMP que algunos autores denominan “Form BIPLLOT” y “BIPLLOT de Covarianzas”, respectivamente (8).

También es posible realizar representación simétrica, tanto por filas como por columnas, mediante la descomposición HJ BIPLLOT (30).

- **BIPLLOT basado en análisis de correspondencia:** Este método es el más versátil basado en descomposición de valores singulares (DVS). Generalmente su aplicación es frente a tabulaciones cruzadas o tablas de contingencia de dos variables categóricas, pero es posible extenderlo a tablas de frecuencias, escalas ordinales, variables binarias y datos codificados (8).

Es posible entender este método como un análisis de perfiles⁴ de filas o columnas de una matriz de datos. Al igual que en ACP dependiendo del tipo de representación que se realice sobre el análisis de correspondencia, es posible obtener dos tipos de BIPLLOT asimétricos que dependerán si se visualizan los perfiles, y su distancia Ji-Cuadrada entre puntos, de las filas o de las columnas (8). También es posible obtener HJ BIPLLOT para representar los perfiles.

Si el análisis realizado, y por ende sus graficas BIPLLOT, se extienden a más de dos variables categóricas se habla de Análisis de Correspondencia Múltiple (ACM). Se puede entender este análisis como todos los posibles AC simples de las subtablas que se pueden obtener del conjunto de variables estudiadas (8,31).

Por su parte, en Ecología es muy común el uso de una variante del AC simple denominado Análisis de Correspondencia Canónico (ACC), en la que se obtienen las dimensiones del subespacio por regresión lineal múltiple (7). Maximiza la relación existente entre un conjunto de variables ambientales y un conjunto de variables de especies (frecuencias o presencia/ausencia), obteniendo la combinación lineal correspondiente (7, 33). Con lo anterior y en conjunto al AC se obtiene una ordenación integrada de las especies y de

⁴ Perfiles: expresión de filas o columnas con relación a totales marginales de las mismas.

los datos ambientales asociados, logrando finalmente la visualización de este análisis mediante gráficos BIPLLOT (7).

- **BIPLLOT de Regresión:** En este tipo de BIPLLOT, se muestran casos (filas) y variables respuesta (columnas) de una matriz de datos, en una representación obtenida de un conjunto de predictores (8).

El enfoque clásico de BIPLLOT no considera una distribución probabilística subyacente, los BIPLLOT de Regresión consideran el hecho de que una aproximación mínimo cuadrática se puede encontrar a través de una descomposición de valores singulares (DVS) y que la distribución de los datos debe ser conocida (32). Con esto se considera la posibilidad de aproximar el BIPLLOT de una matriz a través de ajuste modelo bilineal generalizado, de tipo multiplicativo (30):

$$g(\mu) = AB' \quad (2.3)$$

Siendo g una función de enlace, $\mu = E(Y)$, A y B son matrices de parámetros de rango q y ordenes $(n \times q)$ y $(p \times q)$ respectivamente.

Existen dos tipos de BIPLLOT de regresión dependiente de la función de enlace y la distribución de las variables; en el caso de que la función de enlace sea identidad y las variables son de distribución normal se denominan BIPLLOT de Regresión Lineal; si la distribución es perteneciente a la familia exponencial y la función de enlace es distinta a la identidad se denominan BIPLLOT de Regresión No Lineal (32).

En ocasiones surge la necesidad de agregar variables para efectos sólo de interpretación, esto es, se sabe que las variables no aportan información al análisis a realizar, por ejemplo, no afecta al cálculo de los valores y vectores propios, pero si son relevantes para complementar la interpretación desde un punto de vista conceptual. Este tipo de variables se denominan variables suplementarias o pasivas y son de utilidad para complementar la interpretación de los gráficos BIPLLOT (34, 35).

Acorde a la matriz de datos construida del conjunto de factores de riesgo cardiovascular y determinantes sociales recogidos para cada región de Chile, en esta tesis se utilizarán algunos de los análisis multivariados mencionados anteriormente para describir de la mejor manera posible la asociación existente entre ellos, y su evolución temporal. Representando los resultados mediante gráficos BIPLLOT que correspondan, tomando en cuenta la naturaleza de las variables involucradas, número de individuos y supuestos de cada uno de los métodos expuestos.

3 Capítulo III: Objetivos Generales y Específicos

Objetivo General:

Aplicar técnicas de análisis multivariado (BIPLLOT) que permitan explicar y/o determinar la asociación entre un conjunto de factores de riesgo cardiovascular y

un conjunto de determinantes sociales. Se analizará además si la estructura de asociaciones tiene variación temporal en Chile.

Objetivos Específicos:

- Seleccionar indicadores que reflejen el estado, lo más actual posible, respecto al conjunto de determinantes sociales y factores de riesgo cardiovascular en las regiones de Chile.
- Realizar un análisis descriptivo de la asociación entre cada factor de riesgo cardiovascular con el conjunto de determinantes sociales a escala regional.
- Aplicar análisis multivariado (BIPLOT) considerando múltiples estrategias analíticas que describan, identifiquen y/o expliquen la asociación entre factores de riesgo cardiovascular y determinantes sociales.
- Describir ventajas y desventajas de cada uno de los métodos multivariados utilizados para el análisis de la asociación entre factores de riesgo cardiovascular y determinantes sociales de salud.
- Realizar un análisis descriptivo respecto al cambio temporal de las asociaciones encontradas en la implementación del análisis multivariado (BIPLOT).

4 Capítulo IV: Hipótesis

La presente investigación busca responder a la siguiente interrogante: ¿Cómo se desempeñan los métodos multivariados (BIPLOT) para explicar y/o determinar la asociación entre dos conjuntos de variables como son los factores de riesgo cardiovascular y sus determinantes sociales, a nivel regional en Chile y su variación en el tiempo?

Por lo tanto, la hipótesis es,

Hipótesis de estudio:

El análisis multivariado (BIPLOT) permite representar simultáneamente la asociación existente entre un conjunto de factores de riesgo cardiovascular (outcomes) con un conjunto de determinantes sociales (factores de exposición), y su evolución en el tiempo.

5 Capítulo V: Metodología

5.1 Metodología Estadística

5.1.1 BILOT Clásicos

En general, el método BILOT busca aproximar una matriz de datos X de n observaciones y p variables ($n \times p$), y de rango r , mediante una descomposición de valores singulares (DVS), que se puede escribir como (30):

$$X_{(r)} \cong X_{(q)} = U_{(q)} D_{(q)} V'_{(q)} = \sum_{k=1}^q \alpha_k u_k v'_k \quad (5.1)$$

$$X_{(q)} = A_{(q)} B'_{(q)} \quad (5.2)$$

Donde $X_{(r)}$ es la matriz de datos de orden ($n > p$) y rango r , $X_{(q)}$ es la aproximación de la matriz X de rango q ($q < r$), U y V son matrices singulares ortogonales tales que $U'U = V'V = I$ (I identidad), y D corresponde a la diagonal que contiene los α_k mayores valores singulares. A es la aproximación de rango q de la matriz A ($n \times p$) de marcadores fila y B es la aproximación de rango q de la matriz B ($q \times p$) de marcadores columna.

Así, el BILOT de la matriz X se define sobre el gráfico con marcadores fila $a_i = (a_{i1}, a_{i2})'$ y marcadores columna $b_j = (b_{j1}, b_{j2})'$, tal que para cada i y j se tiene (30):

$$x_{(ij)} = a'_{(i)} b_{(j)} = a_{(i1)} b_{(j1)} + a_{(i2)} b_{(j2)} \quad (5.3)$$

Con la ecuación anterior es posible mostrar en el plano los datos sin tomar en cuenta ninguna distribución de las variables estudiadas, mostrando de manera conjunta

varias variables de la misma forma que un gráfico de dispersión lo hace con dos variables. Si el rango de la matriz de datos X es igual a 2, el gráfico BIPLLOT representa en forma exacta dicha matriz en el plano (30).

Gráficamente un BIPLLOT se ve de la siguiente forma:

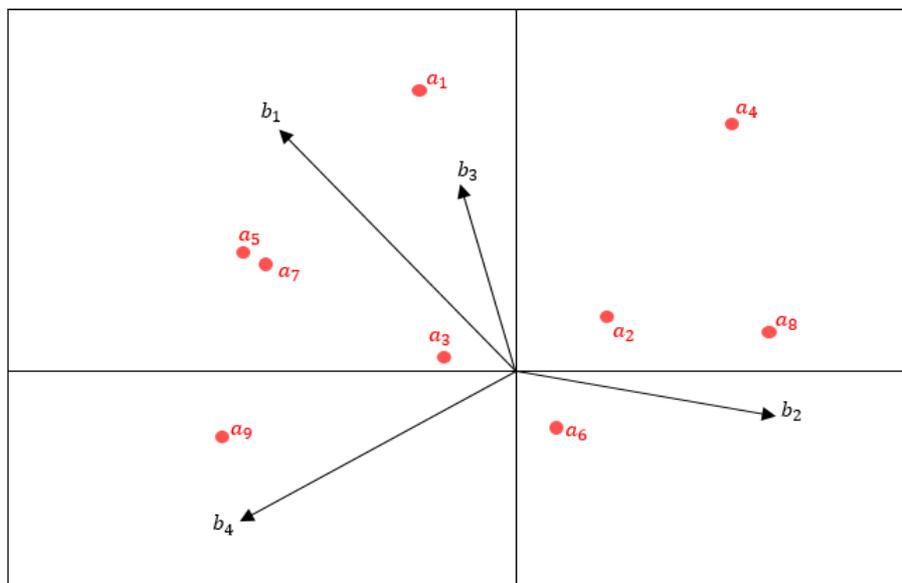


Figura 2. Representación BIPLLOT de una matriz de orden 9x4

Como se aprecia en la Figura 2 las filas a_i de A se pueden ver como puntos y b_j de B como vectores que representan el sentido en que aumentan las variables. Así la interpretación de este tipo de gráficos es muy sencilla; las proyecciones ortogonales de los marcadores filas sobre los marcadores columna aproximan las similitudes de los individuos, el producto escalar entre marcadores columna aproximan la covarianza entre las variables, las longitudes de los marcadores columna aproximan

la desviación estándar de las variables y el coseno del ángulo entre dos marcadores columna aproxima la correlación entre las variables correspondientes (4).

Según Gabriel (25), y en base a la descomposición inicial, esta se puede generalizar como:

$$X_{(q)} = A_{(q)}B'_{(q)} = (UD^s)(D^{1-s}V') \quad (5.4)$$

Y dependiendo de qué valores de s son escogidos, es que las gráficas BIPLLOT tienen distintas propiedades, definiendo los denominados BIPLLOT clásicos. Además, Galindo propone HJ BIPLLOT como alternativa de representación simultanea (36).

Para simplificar la notación en el resto del documento se suprimen los subíndices de las matrices correspondientes.

Si el valor de $s = 1$ en la ecuación 5.4 se encuentra presente un JK BIPLLOT o RMP BIPLLOT (Row Metric Preserving), donde (25, 30):

$$X = AB' = (UD)(V') \quad (5.5)$$

Por lo que se deben escoger marcadores tales que $A = UD$ y $B = V$, que denominados como $J = UD$ y $K = V$ con el fin de representan el nombre de este tipo de gráficos.

Se demuestra que $XX' = JJ'$, por lo que este tipo de gráficos mantiene la métrica para las filas, pero no para las columnas, por lo que se obtiene una alta calidad en la representación para las filas (30).

Cuando el valor de $s = 0$, en la ecuación 5.4 la descomposición queda de la siguiente forma (25, 30):

$$X = AB' = (U)(DV') \quad (5.6)$$

La ecuación 5.6 describe un GH BIPILOT o CMP BIPILOT (Columnn Metric Preserving), se debiera escoger $A = U$ y $B = VD$, o para que tenga sentido con el nombre $G = U$ y $H = VD$.

En este caso se demuestra que $X'X = HH'$, preservando la métrica euclidea usual entre las columnas, pero no entre filas, con lo que se obtiene una calidad mayor a nivel de columnas.

En caso de que la matriz X este centrada por columnas, $X'X$ es proporcional a la matriz de varianzas y covarianzas, tal que $S = X'X/(n - 1)$ encontrándose frente a un BIPILOT de componentes principales, es decir, ajustando los marcadores para que estén directamente relacionados con el Análisis de Componentes Principales (27).

Los SQRT BIPILOT se refieren a aquella factorización de la ecuación 5.7 en la que el valor de $s = 1/2$, resultando $A = UD^{1/2}$ y $B = VD^{1/2}$. En este caso la factorización no es única ya que se puede verificar que $A'A \neq I$ y $B'B \neq I$. La utilización de este tipo de BIPILOT solo se refiere cuando el objeto es aproximar los elementos de la matriz X , como por ejemplo para la diagnosis de modelos en tablas de contingencia, donde tanto filas y columnas son relevantes en el análisis (30).

5.1.2 HJ BIPLLOT

Los HJ BIPLLOT o RCMP (Row Column Metric Preserving) BIPLLOT representan en un mismo sistema de referencia con máxima representación, tanto marcadores fila, como columna (29). La elección de los marcadores en la DVS corresponde a:

$$X = AB' = (UD)(DV') \quad (5.7)$$

Y su factorización es $A = UD$ y $B = VD$, y adoptando la notación $J = UD$ y $H = VD$ se obtiene un BIPLLOT tal que la calidad de la representación tanto para las filas como para las columnas son de la misma calidad obteniéndose una representación simétrica que permite interpretar tanto a los individuos como las variables (30, 36). A desventaja de este tipo de análisis radica en que no permite reproducir los datos originales dada la descomposición de la ecuación 5.7 (4).

Dependiendo del tipo de análisis que se realice con la matriz de datos X se encuentran una enorme cantidad de BIPLLOT diferentes, en esta investigación se estudió el Análisis de Componentes Principales, Análisis de Correspondencia y BIPLLOT de Regresión. Una vez realizados estos análisis es posible utilizar una de las descomposiciones señaladas anteriormente y obtener graficas BIPLLOT.

5.1.3 BIPLLOT Componentes Principales

El Análisis de Componentes Principales se utiliza en caso de variables continuas y su finalidad es reducir el número de variables representando la matriz de datos en una dimensionalidad menor a la original.

En este análisis se aproxima la matriz centrada y/o escalada de rango r a través de su proyección ortogonal de rango menor mediante q vectores lineales e independientes. Estas representan las columnas de una matriz denominada T , tal que $T'T = I$. La solución a este problema radica en encontrar T , simétrica y definida positiva, tal que minimice el error cuadrático medio de las predicciones $\|X_c - \hat{X}_c\|$:

$$X_c \cong X_c T T' = Z T' \quad (5.8)$$

Donde Z contiene a las componentes principales.

Si T es igual a V en la ecuación 5.8 sus columnas coinciden con los autovectores asociados a los q mayores autovalores de la matriz de varianzas covarianzas. Además, si se considera la DVS de la ecuación anterior se obtiene que:

$$X_c \cong (UDV')VV' = UDIV' = UDV' \quad (5.9)$$

Obteniendo la DVS que radica en cada uno de los BIPLLOT clásicos que se grafican en los componentes principales encontrados en el análisis (32).

Luego la varianza explicada por el análisis de componentes principales y la identificación de los componentes se refieren a los mismos de los descritos en el apartado anterior.

5.1.4 BILOT Análisis de Correspondencias

El análisis de correspondencia es una adaptación para datos categóricos del ACP, donde el objetivo es identificar dimensiones que expliquen el máximo de variabilidad de los datos basando su descomposición en matrices centradas o normalizadas utilizando eigenvalores y eigenvectores mediante DVS (37).

Para explicar de manera sencilla este método, el análisis de correspondencia simple basado en tabla de contingencia de 2 columnas y n observaciones es útil.

La matriz definida por estas dos variables en n observaciones se denomina N o de correspondencia. El primer paso es calcular la matriz de correspondencia P , con $p_{ij} = n_{ij}/n$. La suma de las filas y columnas de p corresponde a las frecuencias relativas marginales $r_i = p_{i*}$ y $c_j = p_{*j}$, respectivamente. Estas se denominan masas y se utilizan para normalizar la matriz de correspondencia. Luego bajo la hipótesis de independencia se espera que p_{ij} sea el producto de las masas. La matriz normalizada se denomina S y corresponde a (37):

$$s_{ij} = \frac{p_{ij} - r_i c_j}{\sqrt{r_i c_j}} = D_r^{-1/2} (P - r c') D_c^{-1/2} \quad (5.10)$$

Donde r y c son las masas de filas y columnas, y D_r y D_c corresponden a las matrices diagonales con estas masas en sus respectivas diagonales.

La inercia total se define por la traza de SS' y representa el porcentaje de varianza de la tabla de contingencia. Esta inercia también se puede definir en términos de

Chi cuadrado debido a que los residuos estandarizados de S se asemejan en su forma, por lo tanto se puede definir por χ^2/n (37).

La estructura de asociación en la matriz S , se puede descomponer mediante DVS, como (38):

$$S = UDV' \quad (5.11)$$

Con lo anterior es posible realizar BILOT clásicos basados en esta descomposición.

5.1.5 BILOT de Regresión

En el año 1938, Householder y Young demostraron que la aproximación mínimo cuadrática de una matriz coincide con su descomposición de valores singulares (30), esta premisa permite mostrar esta aproximación mediante graficas BILOT, denominados BILOT de Regresión.

Así, una matriz X se puede ajustar mediante el modelo bilineal general multiplicativo:

$$g(\mu) = AB' \quad (5.12)$$

Siendo g una función de enlace, $\mu = E(Y)$, A y B son matrices de parámetros de rango q y ordenes (nxq) y (pxq) respectivamente. La función de enlace puede tener cualquier distribución de la familia exponencial, tales como la identidad en regresión lineal, o logit, probit, entre otros (32).

Dado que las matrices A y B son desconocidas excepto por su rango, estas deben ser estimadas de manera iterativa por medio de regresiones en forma alternada, esto es; dado A se estima B y dado B se estima A . El problema de esta estimación radica en que puede haber multicolinealidad ya que la solución se basa en el valor inicial de A , esto se puede mejorar utilizando algún método de escalamiento como Componentes Principales (32).

Una vez estimados los valores de A y B se realiza la descomposición DVS con lo cual se obtienen los marcadores filas y columnas permitiendo representar en BIPLLOT ya conocidos (GH, JK, HJ).

5.1.6 Características de los marcadores

Acorde al tipo de BIPLLOT, los marcadores tienen distintas propiedades que pueden llegar a ser muy útiles en la interpretación de las gráficas. Un resumen de estas propiedades se encuentra a continuación (30):

Tipo de Marcadores y Propiedades	Métrica
1. Marcador columna $B = VD$	$A'A = I$
Propiedades: El producto escalar de las columnas de X , coincide con el producto escalar de los marcadores columna, esto es $x_j'x_j \approx b_j'b_j$	

La longitud al cuadrado de los vectores que representan a los marcadores columna, aproximan la covarianza entre las variables correspondientes, de donde: $b_j' b_j = s_{jj} = s_j^2$

La longitud de los marcadores columna, aproxima la desviación estándar de las variables.

El coseno del ángulo entre dos marcadores columna, aproxima la correlación entre variables correspondientes: $\cos(b_j, b_k) \approx \cos(x_j, x_k)$.

Las coordenadas de la matriz de marcadores columna de B en el BILOT de Componentes Principales, son equivalentes a los pesos de las variables a lo largo de los ejes principales.

Es optima la calidad de la representación de las columnas.

2. Marcador columna $B = V$

$$B'B = I$$

Propiedades:

Las coordenadas para las columnas coinciden con las proyecciones de los ejes originales en el espacio de las componentes principales: $I_p V = V = B$.

La similitud entre las columnas, se aproxima utilizando la inversa de la matriz de dispersión de los individuos.

Es baja la calidad de la representación para las columnas.

3. Marcador fila $A = UD$

$$B'B = I$$

Propiedades:

El producto escalar, con la métrica identidad, de las filas de X coincide con el producto escalar entre marcadores filas: $XX' = AA'$

La distancia euclídea entre las filas de X , coincide con la distancia entre los marcadores fila: $(x_i - x_s)'(x_i - x_s) = (a_i - a_s)'(a_i - a_s)$

Los marcadores fila, coinciden con las coordenadas de las filas en el espacio de las componentes principales: $\varphi = XV = A$

Es óptima la calidad de representación de las filas.

4. Marcador columna $A = U$

$$A'A = I$$

Propiedades:

En el BIPLLOT de Componentes Principales, la distancia de Mahalanobis entre dos filas de X , se aproxima mediante la distancia euclídea entre dos marcadores fila, esto es: $(x_i - x_s)'S^{-1}(x_i - x_s) \approx (a_i - a_s)'(a_i - a_s)$

El i -ésimo marcador fila, coincide con las coordenadas del i -ésimo individuo sobre las r primeras componentes principales estandarizadas, esto es:

$$a_i = \psi_i = \sqrt{(n-1)}u_i = y_i'V[\sqrt{(n-1)}D^{-1}]$$

Es baja la calidad de representación para las filas.

Cuadro 1. Propiedades marcadores fila y columna

De lo anterior, y según su descomposición, GH BIPLLOT cumple con las propiedades 1 y 4 reflejando óptima calidad de representación en columnas y no en filas; JK BIPLLOT presenta propiedades 2 y 3 obteniendo óptima calidad de representación

en filas y no columnas; y HJ BIPLLOT presenta las propiedades 1 y 3 logrando óptima representación en filas y columnas.

5.1.7 Calidad de la representación

La bondad de la aproximación de los gráficos BIPLLOT, también llamada calidad global de la representación se define de manera similar al coeficiente de determinación en Análisis de Regresión; como el cociente entre la suma de cuadrados de elementos de X_q y la suma de cuadrados de los elementos de X (30).

De la ecuación 5.4, se puede describir como (27):

$$\left[\frac{\sum_{k=1}^q \alpha_k^2}{\sum_{k=1}^r \alpha_k^2} \right] \quad (5.13)$$

Sin embargo, este indicador no refleja la calidad de la representación individual de filas y columnas que es lo que interesa en este tipo de gráficos. Para mejorar esta evaluación se deben considerar las contribuciones del factor tanto al elemento fila como al elemento columna, permitiendo determinar la parte de la variabilidad de cada elemento explicada por el factor (30).

Así, la contribución del factor al elemento columna es:

$$CFE_{col} = \frac{b_{jk}^2}{\sum_{k=1}^q b_{jk}^2} \quad (5.14)$$

La ecuación 5.14 permite conocer qué variables son las responsables de la ordenación de los individuos. Si los datos están centrados, este valor coincide con

el cuadrado del coeficiente de correlación entre la j-ésima variable y el k-ésimo eje (30).

La contribución del factor al elemento fila es:

$$CFE_{fila} = \frac{a_{ik}^2}{\sum_{k=1}^q a_{ik}^2} \quad (5.15)$$

Este indicador no se puede interpretar en términos de correlación, sino que se interpreta en términos de calidad de representación, mostrando la parte de la información de un individuo recogida por el eje k-ésimo en el plano (30) figura 3.

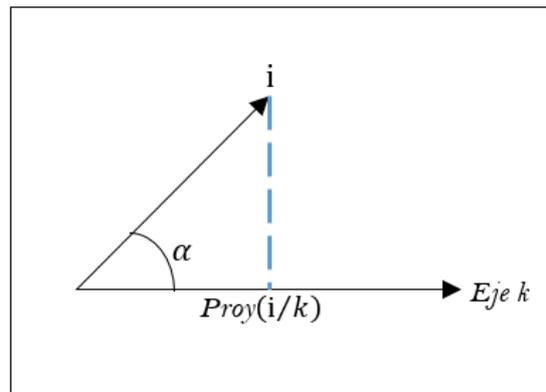


Figura 3. Contribución del Factor al Elemento fila i-ésimo

5.1.8 Variables suplementarias en BIPLLOT

Además, con el fin de mejorar la interpretación se utilizan variables suplementarias. Las variables suplementarias o inactivas, pasivas, externas o adicionales, corresponden a aquellas variables que se excluyen del análisis original, pero luego son incluidas para facilitar la interpretación. La estimación de la posición en el

grafico BIPLLOT de las variables que se incluyen de manera pasiva se realiza mediante mínimos cuadrados. Se supone que se tienen los errores provenientes de la diferencia entre los valores observados y los productos escalares de los datos representados en el BIPLLOT. Las coordenadas de las variables suplementarias definidas por B_v se obtienen a través de los coeficientes de regresión de variables suplementarias Z en los ejes BIPLLOT, esto es, las columnas de B (ecuación 5.5) (34, 35).

$$B'_v = (A'A)^{-1}A'Z \quad (5.16)$$

Con Z que contiene los valores reales de Z y A que corresponde a la matriz que contiene marcadores fila en la DVS que genera el grafico BIPLLOT.

5.2 Metodología del estudio

El presente estudio se refiere a un estudio de tipo Transversal Poblacional, basado en información recogida (cálculo de indicadores) a partir de encuestas poblacionales oficiales en Chile.

5.2.1 Universo y Muestra

La unidad de estudio de esta investigación refiere a las regiones del país, en este caso no existe muestra explícita ya que es posible obtener indicadores para las 15 regiones que representan al país.

Sin embargo, al realizar los cálculos de los indicadores regionales se tomó en cuenta el diseño muestral, ya sea de la Encuesta Nacional de Salud o Encuesta de Caracterización Social. Para el cálculo de cada indicador se utilizaron los ponderadores correspondientes a cada encuesta utilizada.

La Encuesta Nacional de Salud se realizó mediante un diseño muestral complejo, estratificado geográficamente, multietápico y de conglomerados. El marco muestral en el año 2003, proviene de la encuesta ENCAVI del año 2000 con el fin de precisar dicha información (23,24), mientras que, en 2009, el marco muestral se basa en los datos del Censo de Población y Vivienda del año 2002 (39).

Para el año 2003 la muestra tiene una representatividad nacional y urbano rural, por lo que puede que los indicadores regionales no sean del todo robustos, sin embargo, es posible encontrar indicadores a nivel regional. Dado lo anterior se tuvo especial cuidado con los resultados presentados para ese año por región, tomando en cuenta que es solo una referencia. Sin embargo, el año más actualizado y la base de esta investigación se trabajaron con la ENS 2009, la cual además de la representatividad lograda en 2003, se suma el nivel regional⁵.

La tasa de respuesta en la Encuesta Nacional de Salud, fue de un 90.2% en 2003 (23), y un 85% para el 2009 (39).

⁵ <http://epi.minsal.cl/estudios-y-encuestas-poblacionales/encuestas-poblacionales/encuesta-nacional-de-salud/antecedentes-ens>

En el caso de la encuesta CASEN, en el año 2003 se realizó un muestreo estratificado, por conglomerados y polietápico (40) utilizando como marco muestral el Censo de Población y Vivienda del año 2002. En el año 2009, se realizó un diseño muestral estratificado y bietápico, utilizando como marco muestral el resultante del “Programa Integrado de Encuestas de Hogares” (PIDEH) desarrollado por el INE a partir de información cartográfica del Censo de Población y Vivienda del año 2002 (41).

Tanto la versión 2003 como la 2009 tienen una representatividad a nivel nacional, regional y urbano-rural.

5.2.2 Variables o tópicos de estudio. Fuentes de información

Según las fuentes de información disponibles, las variables de estudio utilizadas son las referentes a los siguientes factores de riesgo cardiovascular y determinantes sociales:

- **Factores de riesgo cardiovascular;** Hipertensión, Dislipidemias, Estado Nutricional, Diabetes, tabaquismo, Sedentarismo, Síndrome Metabólico y Riesgo Cardiovascular.
- **Determinantes sociales;** Distribución de ingreso, Alfabetismo, Escolaridad, Cobertura Educativa, Condiciones de vida, Empleo, Ruralidad y Etnia.

Las fuentes de información oficial a nivel poblacional son la encuesta CASEN y ENS para los años 2003 y 2009, a partir de las preguntas asociadas a las variables descritas.

Para efectos de selección de variables de riesgo cardiovascular, la única fuente es la ENS, las variables de estudio son las siguientes:

Variable	Descripción
<i>Hipertensión</i>	
Hipertensión	Prevalencia
PreHipertensión	Prevalencia
<i>Dislipidemias</i>	
Colesterol	Promedio en mg/dl
Triglicéridos	Promedio en mg/dl
<i>Estado nutricional</i>	
Obesidad	Prevalencia
<i>Diabetes Mellitus</i>	
Diabetes Mellitus	Prevalencia
PreDiabetes	Prevalencia
<i>Riesgo Cardiovascular</i>	
Riesgo Cardiovascular	Prevalencia
Otros indicadores	

Tabaquismo	Prevalencia
Sedentarismo	Prevalencia
Síndrome Metabólico	Prevalencia

Cuadro 2. Variables Riesgo Cardiovascular

Respecto a determinantes sociales, se utilizó la información recogida por la encuesta CASEN, aun cuando la información aparezca en la ENS. Esta decisión se debe a los objetivos de cada uno de los instrumentos. Las variables utilizadas de la encuesta CASEN son:

Variable	Descripción
<i>Distribución de ingresos</i>	
Pobreza	Porcentaje de la población en condición de pobreza
Índice 20/20	Índice por región
<i>Nivel educacional y alfabetismo</i>	
Alfabetismo	Porcentaje de la población que sabe leer y/o escribir
Escolaridad	Promedio escolaridad en años
Cobertura Educacional	Porcentaje de cobertura Total (desde preescolar hasta superior)
<i>Condiciones de vida</i>	

Calidad Global de la vivienda	Porcentaje de viviendas con calidad global aceptable.
Desempleo	Tasa de desempleo
<i>Sistemas de salud y otros</i>	
Previsión salud	Porcentaje de beneficiarios sistema público (FONASA)
Ruralidad	Porcentaje viviendas en zona rural
Etnia	Porcentaje de población que pertenece a etnia

Cuadro 3. Variables Determinantes Sociales

La definición de cada una de estas variables se presenta a continuación:

Factores de Riesgo Cardiovascular (ENS 2003, 2009)

- **Hipertensión (2003, 2009):** O prevalencia de Sospecha de Hipertensión, cuando una persona se encuentra en tratamiento farmacológico para hipertensión o el valor de medición de la Presión arterial Sistólica ≥ 140 mmHg o Presión arterial Diastólica ≥ 90 mmHg (23,39).
- **PreHipertensión (2009):** Prevalencia, se define cuando una persona presenta valores de Presión arterial Sistólica entre 120 mmHg y 139 mmHg, o Presión arterial Diastólica entre 80 mmHg y 89 mmHg (21). En este caso se evaluó la pre Hipertensión en menores de 25 años. Esta variable será medida solo el año 2009 debido a que no se tuvo acceso a la base de datos desagregada de la ENS 2003 y corresponde a una variable de elaboración propia.

- **Colesterol (2009,2003):** Promedio regional de niveles de colesterol total medido en mg/dl.
- **Triglicéridos (2009, 2003):** Promedio regional de niveles de triglicéridos medido en mg/dl.
- **Obesidad (2003, 2009):** Prevalencia de Obesidad, definida en la ENS como aquellas personas que presentan IMC ≥ 30 (23,39).
- **Diabetes Mellitus (2003, 2009):** O prevalencia de sospecha de Diabetes, basándose en la definición epidemiológica “glicemia de ayuno ≥ 126 mg/dl” o aquellos sujetos que auto reportan diagnóstico médico de diabetes. Para esta variable se excluyen mujeres embarazadas (23,40).
- **PreDiabetes (2009):** Prevalencia. Se define en todas las personas de 25 años o menos que presentan niveles de glicemia en ayuno entre 100 mg/dl y 125 mg/dl (22). Esta variable será medida solo el año 2009 debido a que no se tuvo acceso a la base de datos desagregada de la ENS 2003 y corresponde a una variable de elaboración propia.
- **Riesgo Cardiovascular (2003, 2009):** Prevalencia de riesgo cardiovascular “Alto” y “Muy Alto”. La ENS 2009 realiza este cálculo cuantitativo mediante el instrumento de National Cholesterol Education Program- Adult Treatment Panel III (ATP III Update, 2004), que calcula riesgo para enfermedades coronarias duras, esto es infarto al miocardio y muerte coronaria (39). En el año 2003 también se utilizó la pauta ATP III (23).

- **Tabaquismo (2003, 2009):** O prevalencia de “Fumador actual”. Se refiere a aquellas personas que fuman actualmente de manera ocasional o a diario (23,34).
- **Sedentarismo (2003, 2009):** O prevalencia “Sedentarismo en tiempo libre”, personas que realizan menos de 30 minutos 3 veces a la semana de actividad física de tiempo libre, es decir fuera del trabajo (23, 39).
- **Síndrome Metabólico (2003, 2009):** Prevalencia de síndrome metabólico calculado según ATP III Update con circunferencia de cintura. Para todas las personas con puntaje mayor o igual a 3 de 5 criterios; circunferencia de cintura, colesterol HDL, triglicéridos, presión arterial y glicemia (23, 39).

Determinantes sociales (CASEN 2003, 2009)

- **Pobreza (2003, 2009):** Porcentaje de hogares en situación de pobreza o de pobreza extrema sobre el de la población. El cálculo de este indicador es través de los ingresos, esto es, hogares que se encuentran bajo la línea de la pobreza y de la pobreza extrema, respectivamente definidas por CASEN⁶.
- **Índice 20/20 (2003, 2009):** Es un índice de desigualdad que muestra la relación del ingreso recibido por el 20% de hogares de mayores ingresos y el correspondiente al 20% de menores ingresos (40, 41).

⁶ Más detalles respecto de línea de la pobreza en http://observatorio.ministeriodesarrollosocial.gob.cl/casen/casen_def_pobreza.php

- **Alfabetismo (2003, 2009):** Se refiere al porcentaje de personas que saben leer y/o escribir calculado en personas de 15 años o más (40, 41).
- **Cobertura Educativa (2003, 2009):** Porcentaje de personas que asisten al sistema educacional entre la población de 4 a 25 años (40, 41).
- **Escolaridad (2003, 2009):** Número promedio de años cursados en el sistema educacional entre las personas de 15 años o más (40, 41).
- **Calidad Global de la Vivienda (2003, 2009):** Porcentaje de viviendas con calidad aceptable en la región (40, 41).

La calidad global de la vivienda se calcula acorde a la siguiente tabla:

Índice de Calidad global de la vivienda			
Tipo de Vivienda	Índice de Saneamiento	Índice de Materialidad	Índice Calidad Global
Aceptable	Aceptable	Aceptable	Vivienda calidad Aceptable
		Recuperable (solo muro recuperable resto aceptable)	
	Deficitario	Aceptable	Vivienda calidad Recuperable
		Recuperable	
Irrecuperable	Aceptable	Aceptable	

		Recuperable	Vivienda calidad Irrecuperable
		Irrecuperable	
	Deficitario	Aceptable	
		Recuperable	
		Irrecuperable	

Cuadro 4. Definición Calidad Global de la Vivienda

Con Tipo de vivienda “Aceptable” se considera a Casa aislada (no pareada); Casa pareada por un lado; casa pareada por ambos lados; Departamento en edificio con y sin ascensor; Pieza en casa antigua o conventillo. Tipo de vivienda “Irrecuperable” se refiere a Mediagua o Mejora; Rancho, choza o ruca; Vivienda precaria; Móvil (por ejemplo, carpa) u Otro tipo.

Por otro lado, el Índice de Saneamiento es “Deficitario” si la disponibilidad agua es deficitaria o el servicio higiénico deficitario y “Aceptable” cuando la disponibilidad agua aceptable y servicio higiénico aceptable.

Finalmente, el Índice de materialidad se define “Aceptable” cuando la materialidad de muros, techo y piso es aceptable, “Recuperable” en el caso de que el muro sea recuperable y un indicador aceptable (piso o techo) o más de un indicador recuperable y ninguno irrecuperable. “Irrecuperable” en el caso de que haya al menos un indicador irrecuperable.

- **Desempleo (2003, 2009):** Porcentaje de la población desocupada (cesantes y personas que buscan trabajo por primera vez) de 15 años y más con respecto a la fuerza de trabajo o población económicamente activa (40, 41).
- **Previsión Salud (2003, 2009):** Porcentajes de personas que están afiliadas a FONASA (40, 41).
- **Ruralidad (2003, 2009):** Porcentaje de personas que viven en zonas rurales en la región (40, 41).
- **Etnia (2003, 2009):** Porcentaje de personas que pertenecen o son descendientes de alguna etnia indígena (40, 41).

5.2.3 Métodos de recolección de la información e instrumentos usados

La información fue recolectada desde los canales oficiales de cada institución:

- Información de ENS, se recogerá de Ministerio de Salud.
- Información CASEN, se recogerá de Ministerio de Desarrollo Social (Ex Mideplan).

Se utilizaron como instrumentos las encuestas poblacionales elaboradas por cada entidad y se establecieron estimadores según los diseños muestrales de cada una de ellas siguiendo los manuales técnicos relativos al diseño muestral.

En el caso del análisis 2003, no fue posible conseguir la base de datos individual que incluya la variable región. Dado lo anterior se decidió utilizar fuentes

secundarias y se levantaron los indicadores acordes al informe de resultados publicado en el Ministerio de Salud (23).

La información recogida más actualizada de las variables de Salud, fue calculada a partir de la ENS 2009. Para ello se calcularon los indicadores según lo descrito en el manual de uso de base de datos de la encuesta (42), y acogido a la corrección de ponderadores y variables descritos en el informe de validación de la base de datos elaborado por el departamento de Epidemiología del MINSAL (43).

Por otro lado, los indicadores relacionados con determinantes sociales fueron calculados directamente de la Encuesta de Caracterización Socioeconómica Nacional (CASEN), tomando en cuenta los manuales técnicos para el usuario y de metodología publicados, tanto para el año 2003 como para el año 2009 (40, 41). Si bien la última versión de la encuesta CASEN es del año 2013, se decidió trabajar con los mismos años que abarca la ENS, con el fin de obtener resultados consistentes respecto a la periodicidad del estudio.

5.2.4 Técnicas de procesamiento de datos

Los cálculos de todos los indicadores regionales utilizados en esta investigación fueron calculados mediante el software estadístico Stata en su versión 13. Tanto la ENS y la encuesta CASEN entregan base de datos en este formato, y se siguieron rigurosamente las instrucciones publicadas en manuales metodológicos de muestreo y de uso de las bases de datos respectivas. Una vez calculados los

indicadores, se procedió a validar los cálculos realizados contrastando con publicaciones oficiales de cada indicador regional a modo de referencia.

Con los indicadores calculados para cada región se aplicaron diversos tipos de análisis estadístico, estos análisis fueron realizados con el software R- Project y MULTBILOT.

R es un lenguaje de programación y entorno de programación utilizado para análisis de datos y gráficos, es un software libre (GNU) desarrollado en un lenguaje similar al lenguaje S, de S-plus⁷. Al descargar R, este programa contiene funciones básicas, para poder ejecutar cálculos más específicos se utilizan distintos paquetes adicionales que se publican en “The Comprehensive R archive network” (CRAN) en <https://cran.r-project.org/>.

Para esta investigación se utilizaron principalmente dos paquetes relacionados con análisis multivariado y BILOT; FactoMineR y dynBiplotGUI.

FactoMineR es un paquete R dedicado al análisis multivariado que tiene la particularidad de utilizar información suplementaria en el análisis y presentar un punto de vista más geométrico de los resultados (44). Tiene funciones de análisis de correspondencia simple y múltiple, y análisis de componentes principales, este último es el que se utilizó en esta investigación como una etapa descriptiva multivariada y de reducción de los conjuntos de variables.

⁷ <https://www.r-project.org/about.html>

Por su parte, dynBiplotGUI⁸ es un paquete R diseñado para el análisis interactivo de gráficos Biplot clásicos y HJ Biplot activando ventanas de fácil utilización y evitando el uso de líneas de comando en el análisis facilitando el manejo de formato de los gráficos BILOT. Este paquete representa claramente las diferencias entre los distintos tipos de BILOT clásicos.

Con el fin de complementar la interpretación de los gráficos BILOT, se decide incorporar a la gráfica las variables no utilizadas en análisis BILOT final mediante variables suplementarias, la función que genere este análisis no fue encontrado en el paquete dynBiplotGUI por lo que se decidió utilizar el software desarrollado por Vicente-Villardón para análisis BILOT llamado MULTBILOT. Este software está desarrollado en base a lenguaje Matlab y es de uso libre, contiene análisis BILOT clásicos, HJ BILOT, Análisis de correspondencia, MANOVA BILOT, BILOT logístico, entre otros análisis. En octubre del presente año se incorporan las variables suplementarias al análisis de este software (45).

⁸ <https://cran.r-project.org/web/packages/dynBiplotGUI/dynBiplotGUI.pdf>

5.2.5 Análisis e interpretación

Los pasos generales del análisis estadístico son los siguientes:

1. Análisis univariado descriptivo de las variables involucradas.

Como en toda investigación es de suma importancia analizar de manera sencilla y descriptiva la información. Para esto se realizó un análisis descriptivo básico. Considerando que el tamaño de muestra es pequeño, se calculó el mínimo y máximo, mediana, y rango intercuartil, cuyo fin principal es apreciar directamente de la base de datos información del panorama del país acorde a las regiones de este estudio.

2. Análisis descriptivo multivariado de las variables.

En BILOT, se supone que la cantidad de variables no debe ser mayor que la cantidad de observaciones en el análisis, esto es, $n \geq p$ en la matriz de datos.

Con el fin de disminuir la cantidad de variables se utilizó Análisis de Componentes Principales observando de manera independiente el conjunto de factores de riesgo y determinantes sociales. Se utilizó este análisis para obtener una mirada descriptiva de ambos grupos de variables y verificar que no haya información redundante, además de cumplir los requisitos para la utilización de métodos BILOT. Las variables no utilizadas fueron representadas en los círculos de correlaciones de manera suplementaria, ya que luego serán recogidas de la misma manera.

Para mayor detalle ver Anexo. Análisis de componentes principales para la reducción de variables.

3. Análisis de asociación factores de riesgo cardiovascular y determinantes sociales

Antes del análisis de asociación con métodos multivariados, se realizó un análisis de asociación simple mediante el coeficiente de correlación de Spearman, considerando que el número de individuos es pequeño. Este análisis solo describe asociación univariada como referencia y no pretende ser utilizado para mayores interpretaciones.

Posteriormente, se realizó el análisis BIPLLOT utilizando los métodos apropiados para las variables activas obtenidas en el punto anterior, considerando los supuestos de los mismos. Además, se incorporaron las variables suplementarias o pasivas en el análisis final.

Se evaluó cual es el que mejor representa la asociación entre el riesgo cardiovascular y los determinantes sociales en las regiones del país, esta evaluación se realizó en base a las siguientes características:

- Pertinencia teórica, y
- Dificultad de la interpretación, buscando aquel método que represente de manera más sencilla la asociación a presentar.

Finalmente se analizaron los cambios de manera descriptiva, entre los años 2003 y 2009, observando cambios en los BIPLLOT obtenidos.

5.2.6 Limitaciones

La principal dificultad de esta investigación radica en el número de individuos, en este caso las regiones del país, ya que limita el número de variables con el que se puede analizar la asociación entre riesgo cardiovascular y determinantes sociales a un máximo de 15 para el 2009 y 13 para el año 2003. Sin embargo, esta limitación también puede ayudar a que la representación sea más sencilla, ya que el gráfico no aparece saturado de información y su interpretación es más clara. Sin embargo, al parecer importante la interpretación de todas las variables incluidas en el estudio, es posible incorporar variables suplementarias para mejorar la interpretación de resultados obteniendo mayor información del problema estudiado.

Lo anterior hace pensar que un estudio de esta índole a nivel comunal para cada región puede entregar información más detallada de las mismas, el problema de este punto es que las fuentes de los indicadores recogidos no tienen representatividad a este nivel y se encuentra con el problema de estimación para unidades más pequeñas.

Por otro lado, no fue posible obtener la base de datos desagregada con la variable región para la ENS 2003, lo que limitó la inclusión de variables de elaboración propia como son la prehipertensión y prediabetes e incorporarlas en la comparación de los años 2003 y 2009.

5.2.7 Aspectos éticos

La información recogida en este proyecto de investigación generó información relevante para el país, por lo que se considera ético establecer conductos de explicitación de la información recogida y de los resultados de esta investigación con el fin que sea un aporte a la Salud Pública de Chile.

Para ello se compromete, una vez aprobada esta tesis, al menos:

- Publicar los resultados de este análisis en algún congreso, revista, etc.
- Enviar un resumen de resultados a entidades gubernamentales involucradas, esto es Ministerio de Salud y Ministerio de Desarrollo Social.

6 Capítulo VI: Resultados

6.1 Análisis descriptivos y reducción de variables

Factores de riesgo cardiovascular

Una primera mirada a los datos refleja que el porcentaje de sospecha de hipertensión y de prehipertensión en el país se concentra principalmente en la zona centro sur. A nivel nacional la hipertensión y prehipertensión en menores de 25 años presentan valores de 26,66% y 32,53%, respectivamente. En ambas variables los niveles más altos se ven hacia la zona sur, llamando la atención que la prehipertensión llega a valores sobre 40%, con 40,3% en la VII Región y 43,88% en la VI Región (en hipertensión la VI región llega al máximo, con 33,93%), alejado de valores nacionales. Si bien la prehipertensión parece ser más alta a nivel nacional, la dispersión regional es más amplia, ya que el valor mínimo (6,27% en la región XV) se aleja bastante del valor nacional, esto no ocurre en el caso de hipertensión, donde nuevamente la XV presenta el mínimo, pero con 17,37%.

Las variables relacionadas con dislipidemias, colesterol y triglicéridos, no presentan fuertes variaciones entre las regiones del país. El valor mínimo de colesterol promedio se encuentra en la VI Región (180,4 mg/dl) y el máximo en la región XII (200,07mg/dl), esto se puede deber a que en el sur extremo del país la alimentación tiene mayor contenido graso. Llama la atención que el valor máximo del promedio de triglicéridos es en la II Región, con un valor de 180,38 mg/dl, mientras que el

valor mínimo es de 120,54 mg/dl en la VI Región contrario a lo ocurrido con la hipertensión en esta región. Los valores nacionales no se escapan de lo descrito anteriormente, con valores de 189,12 mg/dl en colesterol y 143,68 en triglicéridos.

El porcentaje de personas con sospecha de diabetes va de 5,28% en la X Región a 13,05% en la VI Región, destacando que de la X-XII región se encuentran valores menores al 6%, y que los más altos porcentajes se encuentran en zonas centrales, VI-VII Región, con valores sobre el 10% de la población. Respecto a la prediabetes en menores de 25 años, los resultados son muy bajos (0% en 5 regiones) resultando llamativo que la región de Valparaíso tenga índices de 8,15%.

La Obesidad se encuentra entre 17,4% y 35,2% en las I y XI Región, respectivamente, nuevamente destacando la zona centro sur en el país, donde se encuentran porcentajes mayores a 30%, cabe destacar que la tercera región también se encuentra sobre este rango.

A nivel nacional el sedentarismo es importante, ya que un 89% declara no hacer deporte en su tiempo libre, la región que posee el valor más bajo en esta variable es la XII (80,77%), y la más alta es la primera región (94,5%). Lo anterior sumado a que el porcentaje de personas fumadoras actuales llega a 40,8% (25,9% en XIV Región y 50,13% en XI Región) en el país es realmente preocupante.

Respecto a la prevalencia de síndrome metabólico y riesgo cardiovascular, se puede decir que a nivel nacional llegan a un 28,10% y 15,37% respectivamente, lo cual parece lógico considerando los valores de hipertensión, obesidad, diabetes,

colesterol, triglicéridos vistos anteriormente. La región que posee menor cantidad de personas con síndrome metabólico es la XII (22,91%) y la mayor con 37,5% es la III, mientras que con respecto a riesgo cardiovascular la X región (8,5%) posee menor porcentaje y la IX mayor (20,6%). Por otro lado, se decidió realizar análisis de componentes principales con el fin de observar las variables de salud, y su correlación con los primeros dos factores pensando en la representación BIPLLOT posterior.

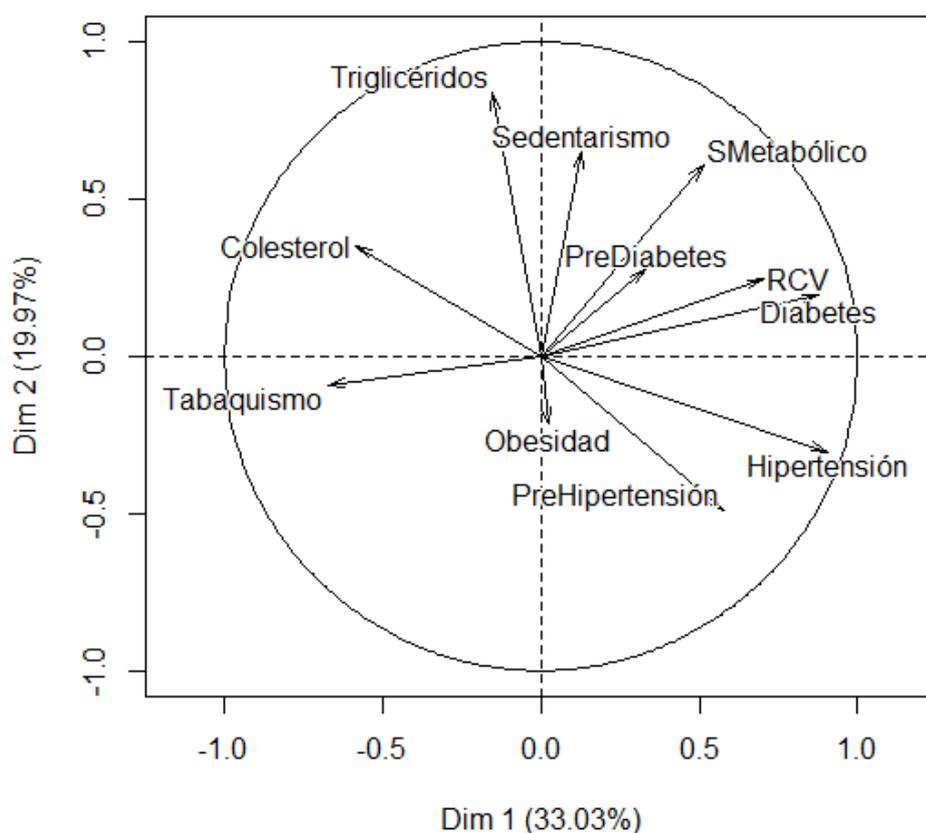


Figura 4. ACP Variables Riesgo Cardiovascular.

Como se ve en la figura 4, los primeros dos factores explican 53% de la varianza total. Este valor no es bueno, por lo que resulta razonable observar que sucede con las variables de salud. Se aprecia que las variables síndrome metabólico y prediabetes aportan correlación a los factores en la misma dirección, esto puede deberse a que el síndrome metabólico es calculado mediante factores de riesgo y uno de ellos está relacionado con la diabetes (glicemia), dado lo anterior se decide mantener como variable pasiva a la pre diabetes considerando que su aporte a los factores es menor. Lo mismo ocurre para las variables riesgo cardiovascular y diabetes, que nuevamente se puede deber a que el indicador de riesgo cardiovascular tiene como componente la diabetes, en este caso y dado el aporte a los factores se decide utilizar variable de diabetes como variable activa y el riesgo cardiovascular como variable pasiva. Finalmente, dada la baja correlación entre obesidad y los otros factores estudiados, se decide dejar esta última como variable pasiva.

Con lo anterior, se realizó un ACP tomando en cuenta las variables obesidad, pre diabetes y riesgo cardiovascular como variables suplementarias o pasivas.

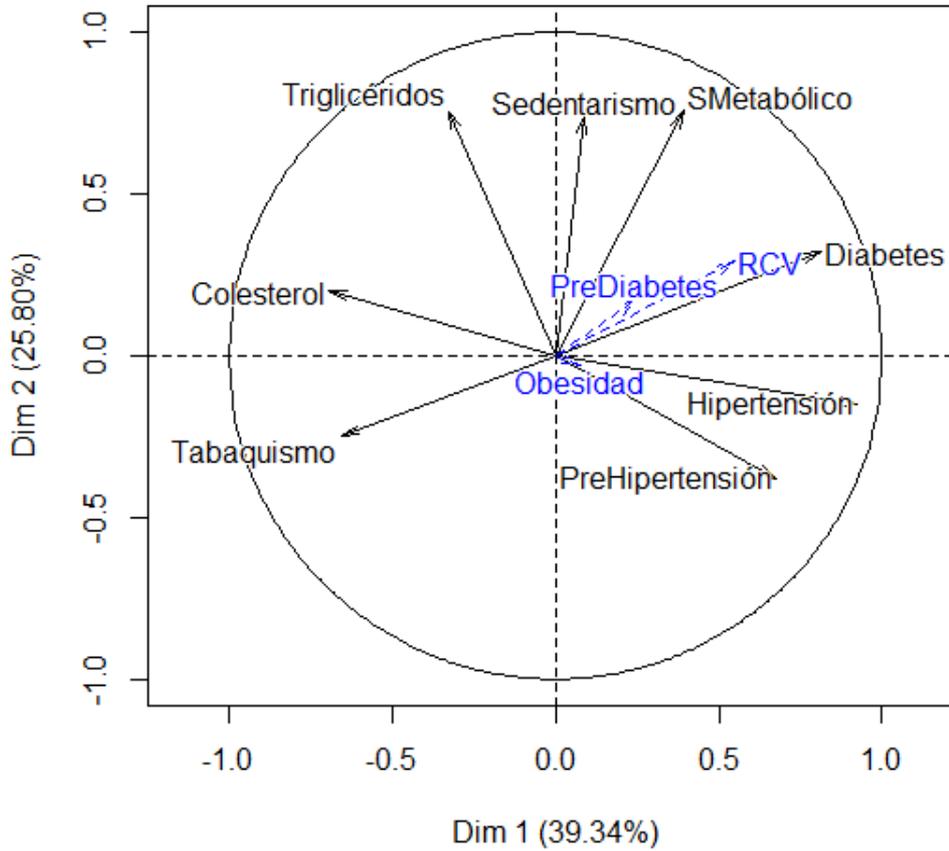


Figura 5. ACP Variables Riesgo Cardiovascular con variables suplementarias
 La figura 5 presenta una gran mejora respecto a la varianza explicada, logrando un 65.14%, parece razonable mantener las variables activas descritas como factores de riesgo cardiovascular para el análisis de asociación mediante BIPLLOT.

Determinantes sociales

La desigualdad regional es clara en nuestro país, en general la zona norte del país se ve favorecida en variables relacionada con ingresos que puede explicarse por la industria minera, mientras que la ruralidad del sur se ve reflejada en indicadores menos favorables respecto a estas variables. En el caso de la pobreza, la II región presenta un porcentaje menor (8.01%), mientras que la zona centro sur aumenta el porcentaje de personas en esta situación, destacando la novena región con un índice mucho mayor, llegando a 27,1%. Al analizar el índice 20/20, ocurre exactamente lo mismo, en la zona norte la I y II región presentan índices alrededor de 7,2 y la IX Región 28,48.

Otra variable que se observa con diferencias es el desempleo, donde en zonas mineras, norte de Chile, es bastante menor, llegando a 5,8 en la II Región. La Región Metropolitana ya presenta niveles de desempleo sobre los 10 y hacia la zona centro-sur se mantiene en esos valores, siendo la novena región la de mayor desempleo.

Las variables asociadas con educación de los habitantes de cada región presentan diferencias menos amplias que aquellas relacionadas con ingresos y empleo. Tanto el nivel de escolaridad y alfabetismo es muy similar a lo largo del país, obteniendo que más del 90% sabe leer y/o escribir (92,27% VII región y 98,68% XV Región) y la escolaridad promedio regional supera los 9 años (9 años en VII región y 11,3 en RM). La cobertura educacional en población de 4 a 25 años, existe una diferencia

levemente más notoria, donde rango va desde 69,1% en la VI región y 75,1% en la VIII.

Los más altos niveles de calidad de la vivienda, considerando tipo de vivienda, materialidad y saneamiento, se encuentran en la RM (67,1% aceptable) y hacia el sur. Esto se puede explicar por el acceso a materiales y la modernización de la ciudad, y hacia al sur el clima que conlleva a que las viviendas deban ser de mejor calidad que en regiones en que las inclemencias del tiempo no afectan durante tanto tiempo en el año, logrando índices sobre el 80% aceptable en el sur y entre un 70% y 80% en el norte.

La cobertura del Fondo Nacional de Salud (FONASA), varía entre regiones, donde los valores más bajos se encuentran en la Región Metropolitana y la II Región con una cobertura de 71,4% y 71,74%, respectivamente. De la Tercera región hasta la IV región (sin considerar RM), los niveles se encuentran sobre un 80%.

Respecto al porcentaje de viviendas de la región que se encuentran en zona rural, es una de las variables con mayor rango, llegando a 24%. Llama la atención que la región con menos porcentaje de ruralidad no sea la RM (2,87%), sino que la II (1,7%) región, ya que se esperaría lo contrario, el crecimiento económico por la minería puede explicar esto. Las regiones con mayor porcentaje en zona rural son mayoritariamente en la zona sur desde la VI región a la de Los Ríos que presentan entre un 29%-32% de viviendas en zonas rurales, excepto la VIII región que tiene solo un 16,7%.

Finalmente, el porcentaje de personas que pertenecen a alguna etnia o son descendientes, la región de la Araucanía alcanza un 30,08%, siendo el porcentaje más alto del país, donde la etnia predominante es la Mapuche. También resalta la zona norte del país, donde en la XV Región y la I Región, hay 25,43% y 12,6%, respectivamente, siendo esta en su mayoría población Aymara. En la zona central los niveles de porcentaje de población étnica baja, llegando a niveles entre 1,6% y 4,08%, en las VII región y la Región Metropolitana.

En primera instancia se decide no utilizar la pobreza dentro del análisis de variables activas en los componentes principales porque por definición el índice 20/20 incluye esta característica y parece medir de mejor manera la desigualdad. Sin embargo, parece de suma importancia incorporarla en el análisis de variables suplementarias. En el caso de los determinantes sociales, la varianza explicada por los dos primeros componentes es de 72% (figura 6). A pesar de que este valor es suficiente, en el círculo de correlaciones se aprecia que algunas variables aportan de manera similar en el plano y se refieren conceptualmente a lo mismo. Con el fin de obtener un número reducido de variables para cumplir con los supuestos del análisis BIPLLOT y no tener información redundante, es posible utilizar algunas variables como suplementarias.

Las variables escolaridad y alfabetismo presentan este comportamiento, y como ya se apreció en el análisis descriptivo no presentaban diferencias amplias a lo largo del país, por lo que se decide utilizar la variable escolaridad como suplementaria ya

que presenta en menor grado correlación con los factores. Por otro lado, el desempleo y el índice 20/20 están en la misma situación anterior y dado que son variables que tienen relación directa con ingreso, se decide utilizar desempleo como variable suplementaria mejorando el porcentaje de varianza explicada. Finalmente, Previsión definido como el porcentaje de personas que pertenece a FONASA y Ruralidad aportan casi de la misma manera, dado que previsión aporta muy poco menos se decide dejar también como variable pasiva.

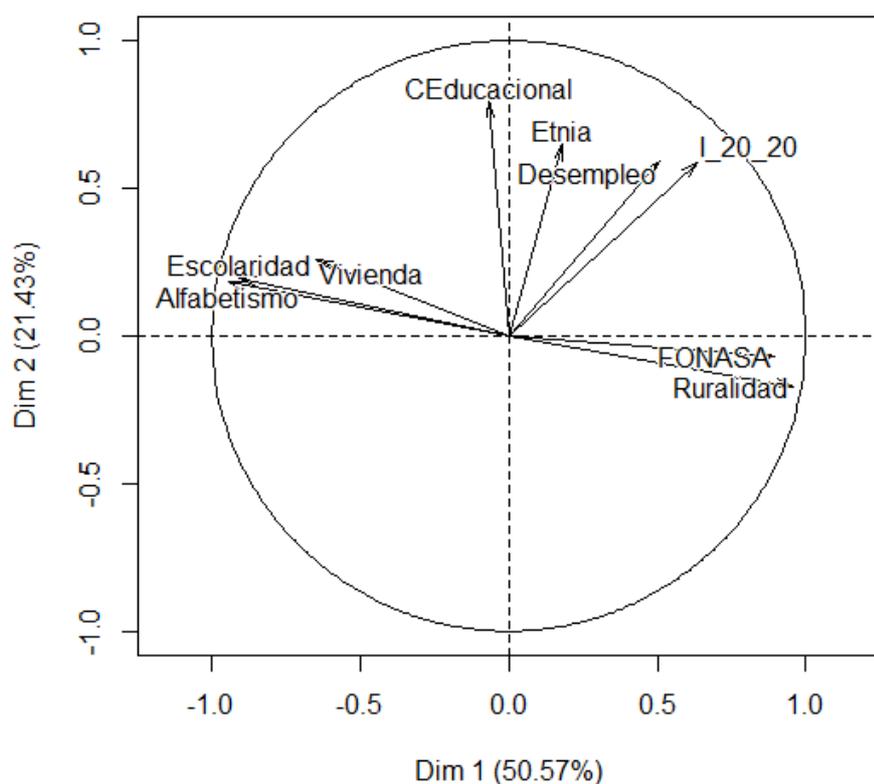


Figura 6. ACP Variables Determinantes Sociales

Así, en el segundo análisis de componentes principales la varianza total explicada aumenta a un 74,18% (figura 7), mejorando la representación y disminuyendo la cantidad de variables activas a utilizar en el análisis.

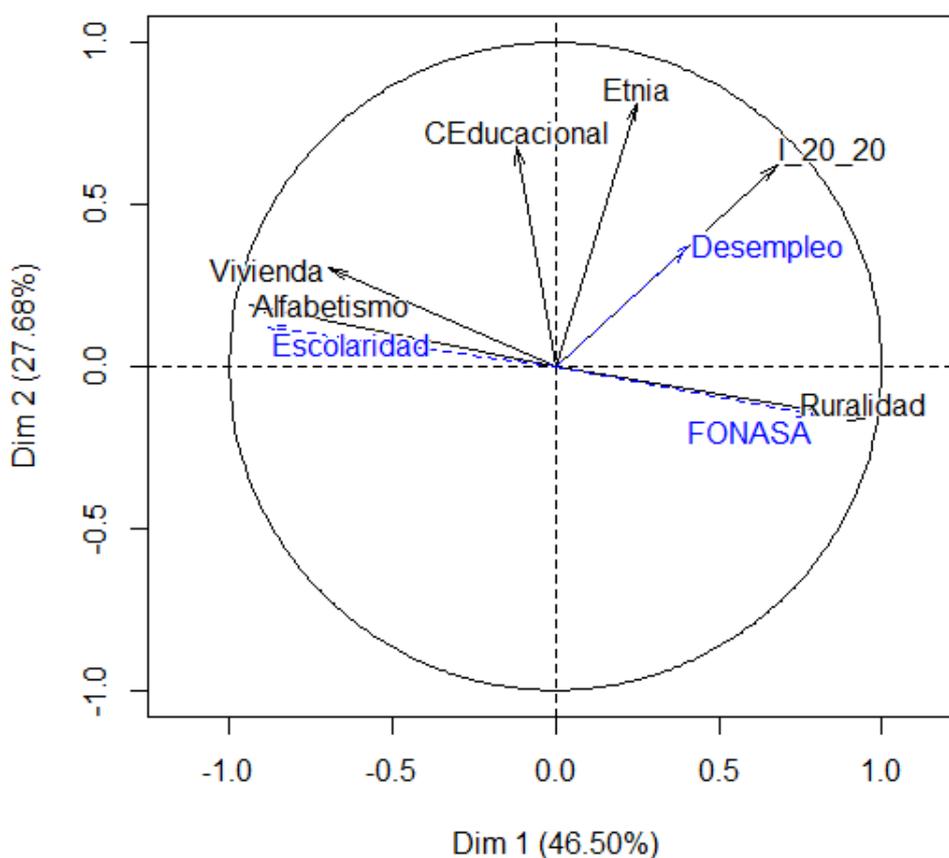


Figura 7. ACP Variables Determinantes Sociales con Variables Suplementarias
Un resumen de las variables a utilizar en el análisis final de esta investigación con el fin de ver la asociación entre riesgo cardiovascular y determinantes sociales son:

Variables Activas	Variables Suplementarias
<i>Riesgo cardiovascular</i>	
Triglicéridos Colesterol Tabaquismo Sedentarismo Síndrome Metabólico Diabetes Sospecha Hipertensión Pre Hipertensión	Riesgo cardiovascular ATP III Update Pre Diabetes Obesidad
<i>Determinantes sociales</i>	
Alfabetismo Calidad Global de la Vivienda Cobertura Educativa Índice 20/20 Ruralidad Etnia	Escolaridad Desempleo FONASA Pobreza

Cuadro 5. Definición variables activas y pasivas para análisis BIPLLOT

6.2 Asociación factores riesgo ECV y determinantes sociales

Primero se realizó un análisis descriptivo de la asociación existente entre las variables estudiadas, utilizando el coeficiente de correlación de Spearman para las variables activas de riesgo cardiovascular y su relación con los determinantes sociales.

En general la asociación univariada no parece ser muy alta a pesar de estar sobre 0,5; tampoco se aprecian asociaciones significativas con un nivel de significancia de 0,05. Sin embargo, la sospecha de hipertensión presenta correlación inversa

significativa con alfabetismo (-0,53), esto refleja que en regiones con niveles de alfabetismo alto disminuye la prevalencia de sospecha de hipertensión. El colesterol se asocia directamente con etnia (0,68), mientras que los promedios regionales de triglicéridos se relacionen directamente con variables como cobertura educacional (0,61) y alfabetismo (0,56), esta última presenta una asociación similar con tabaquismo (0,55). Por su parte, llama la atención que el ser fumador actual también se relaciona directa y altamente con el porcentaje de viviendas de calidad global aceptable (0,7), siendo una de las asociaciones más altas. En regiones donde el porcentaje de viviendas con calidad global aceptable es menor se ve que hay un aumento en la prevalencia de diabetes, alcanzando a -0,75 su coeficiente de correlación. Finalmente, el síndrome metabólico también se relaciona de manera inversa con el porcentaje de viviendas con calidad global aceptable (-0,52).

Con lo anterior se reportó el comportamiento de la asociación univariada entre factores de riesgo cardiovascular y determinantes sociales, sin embargo, una mirada multivariada puede explicar mejor el comportamiento descrito tomando en cuenta las variables en su conjunto y el comportamiento de las regiones respecto a las mismas.

Para seleccionar el tipo de análisis BIPLLOT a realizar, se tomó en cuenta la pertinencia teórica y su facilidad en la interpretación. Dado lo anterior, los métodos BIPLLOT SQRT no serán abordados en esta investigación, ya que se utilizan mayormente en diagnosis de modelos y en matrices construidas a partir de tablas

de contingencia. Los métodos BIPLLOT para el análisis de correspondencia se utilizan cuando las matrices de datos corresponden a datos de conteo provenientes de variables de tipo cualitativa, dada la naturaleza continua de los datos en la presente investigación este método no se utiliza. Por su parte los BIPLLOT de Regresión se basan en la premisa de que la matriz de datos presenta una distribución normal multivariada u otra, lo que no ocurre en este caso. Además, considerando que el número de individuos es reducido no pareció pertinente el uso de este análisis.

Así, para efectos de este análisis se realizaron gráficos BIPLLOT clásicos basados en componentes principales, esto es GH BIPLLOT, JK BIPLLOT y HJ BIPLLOT. Esta decisión se basa en que las representaciones de los BIPLLOT clásicos su grafica tomando como referencia los Componentes Principales del espacio y que la naturaleza de las variables es cuantitativa. Todos los análisis se realizaron en base a la transformación de las matrices de datos, centrando; restando la media y escalando; dividiendo por su la desviación estándar. Para el caso de los BIPLLOT clásicos, se obtuvieron graficas basadas

En base al análisis de componentes principales se ve que al utilizar los dos primeros factores se explica solamente un 55,14% de la varianza total, dado esto y que al utilizar 3 factores es posible obtener una varianza explicada acumulada de 71,38%, por lo que se decidió utilizar los primeros 3 factores.

En base al ACP se realizó la descomposición de valores singulares, obteniéndose BIPLLOT GH, JK y HJ con el fin de ver las diferencias entre cada uno de ellos. Para tales efectos solo se presentan a continuación las gráficas en el espacio constituido por los primeros 2 factores.

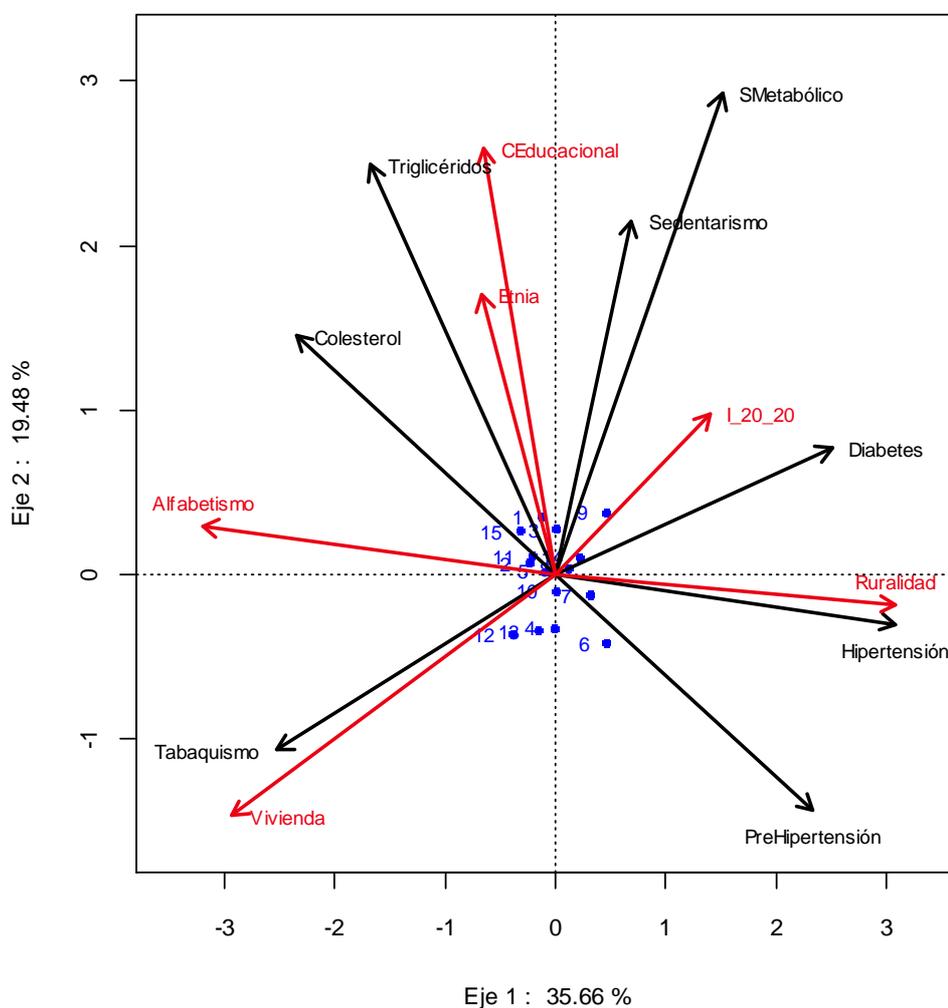


Figura 8. Plano 1-2. GH BIPLLOT 2009

Tal como se esperaba con GH BIPLLOT se obtiene una representación basada marcadores columna, mostrando de manera principal el resultado para las variables y en segundo plano a las regiones. Esto refleja la desventaja principal de este tipo de gráficos, ya que no representa con alta calidad las filas u observaciones.

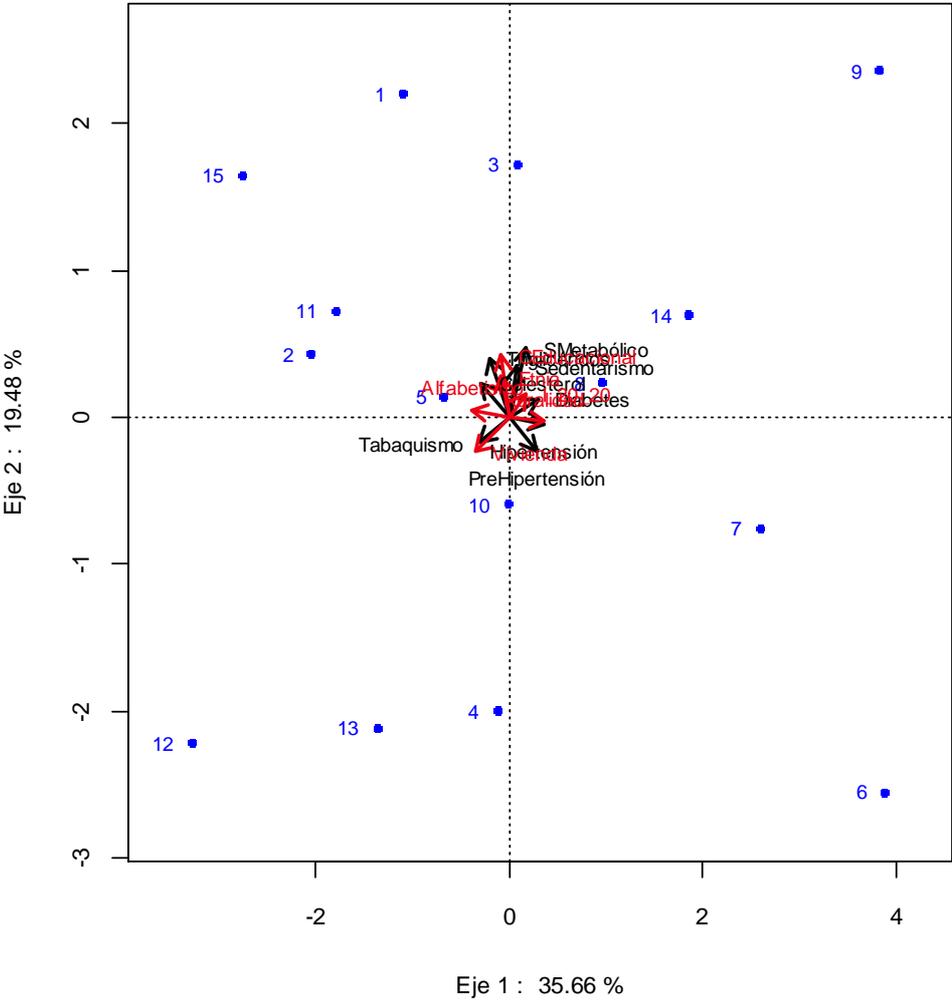


Figura 9. Plano 1-2. JK BIPLLOT 2009

Por otro lado, la gráfica JK BIPLLOT ofrece mejor calidad de representación para las regiones (filas) que, para las variables, por lo que su desventaja radica en que la calidad de la representación de las variables no es del todo buena. Además, no cumple con los requerimientos para esta investigación, ya que se desea interpretación clara de ambas.

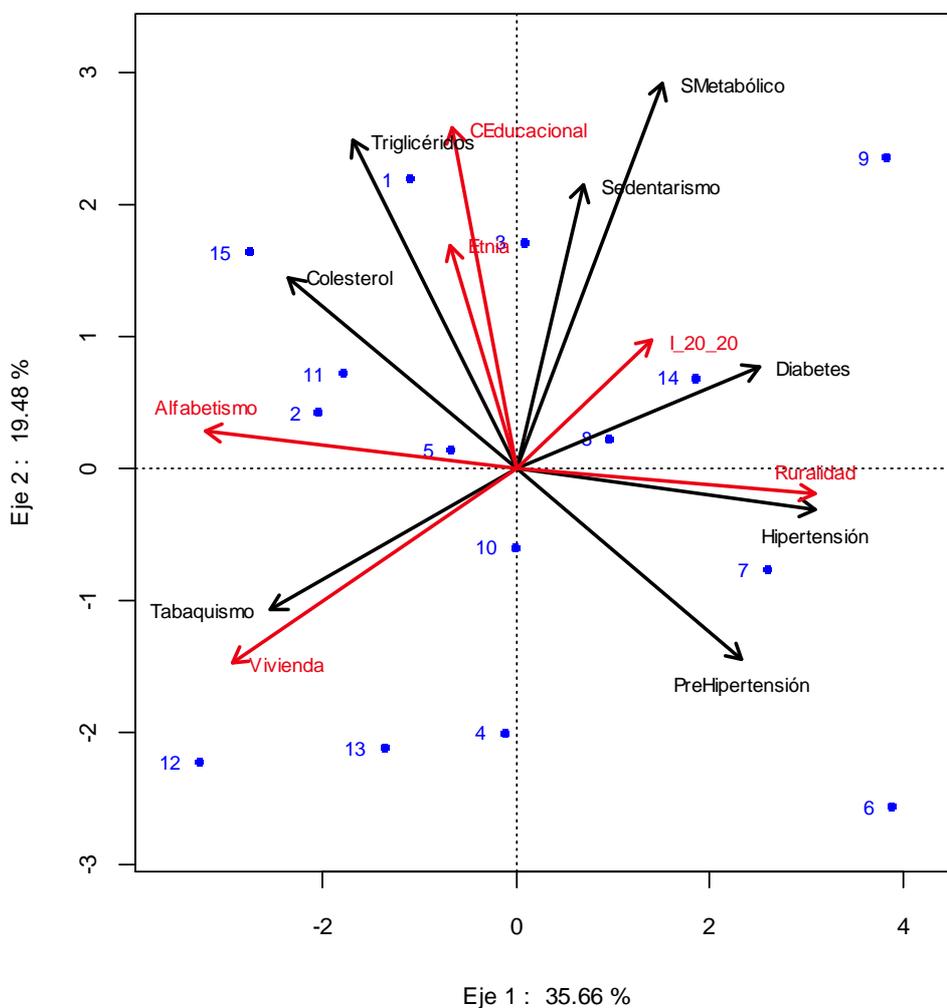


Figura 10. Plano 1-2. HJ BIPLLOT 2009

Parece razonable mantener esta última grafica para la interpretación del análisis multivariado, ya que representa en “primer plano” tanto a las variables como a los individuos, característica principal de los gráficos HJ BILOT. Como ya se mencionó, este tipo de gráficos obtiene óptimas representaciones tanto de filas, como de columnas.

Tal como se mencionó anteriormente se decidió mantener las variables apartadas en el análisis descriptivo de componentes principales como variables suplementarias. Dado lo anterior se realiza un HJ BILOT con variables suplementarias mediante el software MULTBILOT.

MULTBILOT entrega las contribuciones del factor a elemento en una escala de 0 a 1000. Como en este caso se tienen tres factores, el factor en que sea más alta la contribución a las columnas es donde se debe interpretar la variable correspondiente, y en el factor en que más contribuya a la fila corresponderá interpretar la región. Para definir en qué pares de factores se debe interpretar de mejor manera, tanto a las variables como a las regiones, se debe observar las dos contribuciones más altas.

Las contribuciones relativas al factor permiten explicar la contribución de variabilidad de cada variable al factor, en este caso se observó que la mayoría de las variables quedan mejor representadas en el primer factor (cuadro 6) con un total de 8 variables, mientras que en el segundo factor se encuentran 2 variables y en el tercer factor solo dos. Las variables hipertensión, prehipertensión, tabaquismo, síndrome

metabólico, cobertura educacional, calidad global de la vivienda, pueden interpretarse con buena calidad de representación en el plano 1-2. Además, la variable triglicéridos también se interpreta en el plano 1-2, aunque la segunda contribución más alta este en el tercer factor, la diferencia con el primero es mínima. Por lo tanto, en el plano 1-3 corresponde interpretar las variables colesterol, diabetes, ruralidad y alfabetismo.

Respecto a las variables sedentarismo y etnia, presentan una contribución más alta en el plano 2-3, pero con el fin de simplificar la interpretación y no agregar un gráfico extra para estas dos variables se decide mantenerlas en el plano 1-2 y 1-3, respectivamente.

Contribuciones			
Variable	Factor 1	Factor 2	Factor 3
Sospecha de hipertensión	680	7	0
Prehipertensión	389	147	11
Colesterol	393	151	226
Triglicéridos	201	443	219
Diabetes	449	42	243
Tabaquismo	455	81	0
Sedentarismo	34	330	145
Síndrome Metabólico	164	611	1

Cobertura Educacional	30	478	1
Calidad Global de la Vivienda	613	154	42
Ruralidad	681	2	154
Etnia	32	206	603
Alfabetismo	731	6	118
Índice 20/20	140	68	511

Cuadro 6. Contribuciones del factor al elemento columna análisis 2009.

En cuanto a la contribución de las filas, en este caso de las regiones, la mayoría se ven representadas en el plano 1-2, exceptuando las regiones II, V, X y XI, que se ven mejor representadas en el plano 1-3 (cuadro 7).

Contribuciones			
Región	Factor 1	Factor 2	Factor 3
I	97	394	82
II	380	17	500
III	1	279	231
IV	2	473	99
V	74	4	274
VI	615	267	53
VII	555	47	22
VIII	133	8	5

IX	543	209	142
X	0	41	507
XI	272	45	355
XII	441	201	211
RM	178	421	10
XIV	402	57	139
XV	561	201	2

Cuadro 7. Contribuciones del factor al elemento fila análisis 2009.

En la figura 11 se muestra el grafico HJ BIPLLOT de los conjuntos de variables; riesgo cardiovascular (color negro) y determinantes sociales (color rojo) representadas como vectores, y las observaciones; regiones que se encuentran representadas como puntos en color azul oscuro. Además, las variables suplementarias se representan con color azul.

Se aprecia una relación casi nula entre calidad global de la vivienda y prehipertensión (ángulo casi de 90°), por otro lado, se observa una asociación inversa con hipertensión lo que significa que mientras menor sea la calidad de la vivienda mayores tasas de hipertensión, esto mismo ocurre al observar la relación de la calidad de la vivienda con síndrome metabólico y sedentarismo. Lo anterior puede suceder debido a los ingresos, ya que el índice 20/20 también se asocia, pero directamente, con estos factores de riesgo. Llama la atención la asociación directa entre síndrome metabólico, sedentarismo y triglicéridos, y cobertura educacional, ya

que a mayor cobertura educacional se encuentran mayores niveles de riesgo y lo que se esperaría es que con un mayor capital cultural en las regiones disminuyan estos riesgos. Dado lo anterior la sociabilización de información respecto al riesgo cardiovascular en instancias educacionales podrían ayudar en la disminución de estas tasas, algo que ya parece ocurrir respecto a la prehipertensión en donde coberturas educacionales en las regiones altas reflejan menores prevalencias de prehipertensión en menores de 25 años.

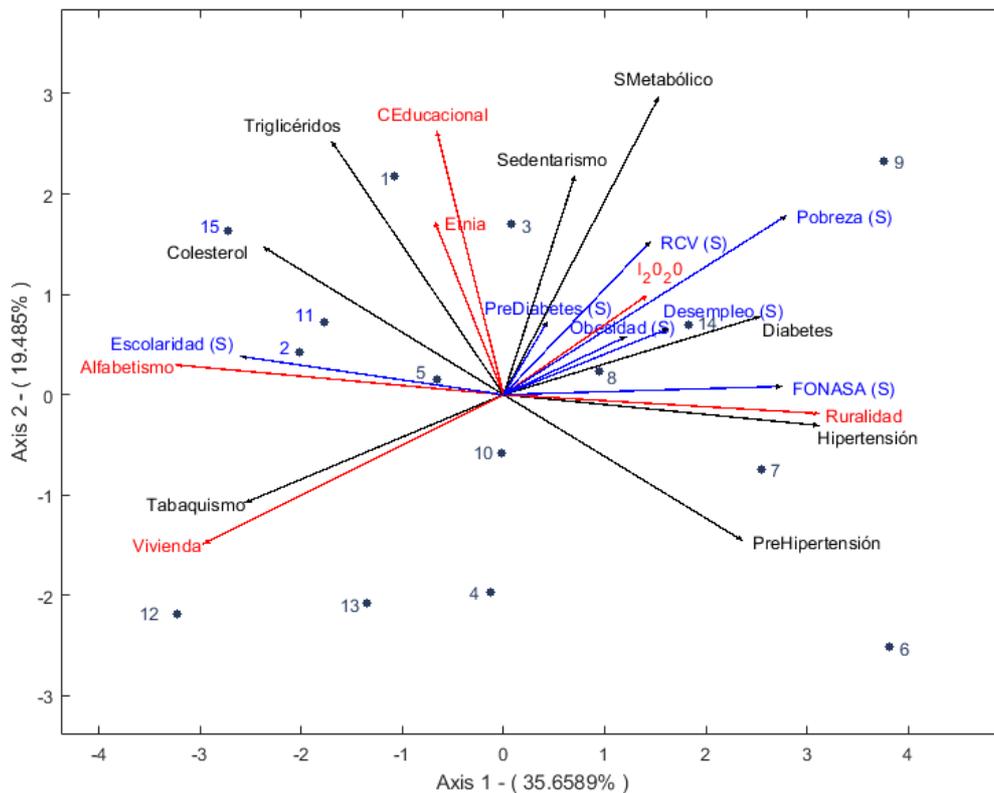


Figura 11. Plano 1-2. HJ BIPLLOT 2009 con Variables Suplementarias

El tabaquismo se asocia fuertemente con la calidad global de la vivienda, lo que refiere a que mejor calidad significa mayor porcentaje de fumadores en la región. Además, es posible apreciar que a mayor cobertura educacional en la región menores serán las tasas de tabaquismo de la misma.

Respecto a las variables suplementarias se aprecia que la pobreza tiene una estrecha relación con el índice 20/20, desempleo, obesidad y riesgo cardiovascular, además representa muy bien a la IX región. Por otro lado, la Hipertensión tiene una relación estrecha con el porcentaje de habitantes que pertenece a FONASA y lo mismo ocurre entre. La escolaridad se relaciona directamente con el promedio del nivel de triglicéridos y con tabaquismo, lo que indica que a mayor cantidad de años de estudio mayores niveles de triglicéridos y tabaquismo.

Respecto a las regiones, es posible notar que aquellas que se pueden interpretar en el plano 1-2 no se encuentran dispersas en el plano.

Las regiones IV, X, XII y RM, presentan niveles de tabaquismo alto y por ende calidad de la vivienda elevada. Tal como se apreció en el análisis descriptivo, la IX región presenta elevados índices de sedentarismo y síndrome metabólico y la desigualdad de los ingresos es alta, lo mismo ocurre, pero en un grado menor en las VIII y XIV región. La séptima región se encuentra aislada respecto de las otras regiones caracterizada principalmente por las variables relacionadas con hipertensión. Por su parte la tercera región se caracteriza principalmente por el

sedentarismo, síndrome metabólico, y junto a las XV y I con respecto a triglicéridos y cobertura educacional.

Respecto al plano 1-3 acumula un 51,9% de variabilidad. En la figura 12 resalta la asociación directa entre el porcentaje de habitantes de las regiones que pertenecen a una etnia y los niveles promedio de colesterol, así como la asociación inversa y fuerte de estos con diabetes. Por su parte existe una asociación baja entre alfabetismo, con diabetes y colesterol, ocurriendo lo mismo con ruralidad.

Las II y la V región se ven descritas por variables como triglicéridos⁹ y alfabetismo, mientras que la X y XI región se caracterizan por niveles de colesterol y etnia elevados.

Tal como ya se conoce la obesidad (suplementaria) presenta correlación directa con colesterol, sin embargo, se destaca que la variable etnia se encuentre fuertemente relacionada también. Por otro lado, las variables pasivas RCV y Prediabetes están asociada con diabetes, también aportan a esta última, pero en menor grado; pobreza, FONASA y escolaridad. Existe una asociación fuerte de estas últimas variables con la ruralidad.

⁹ Con una contribución menor respecto a las columnas.

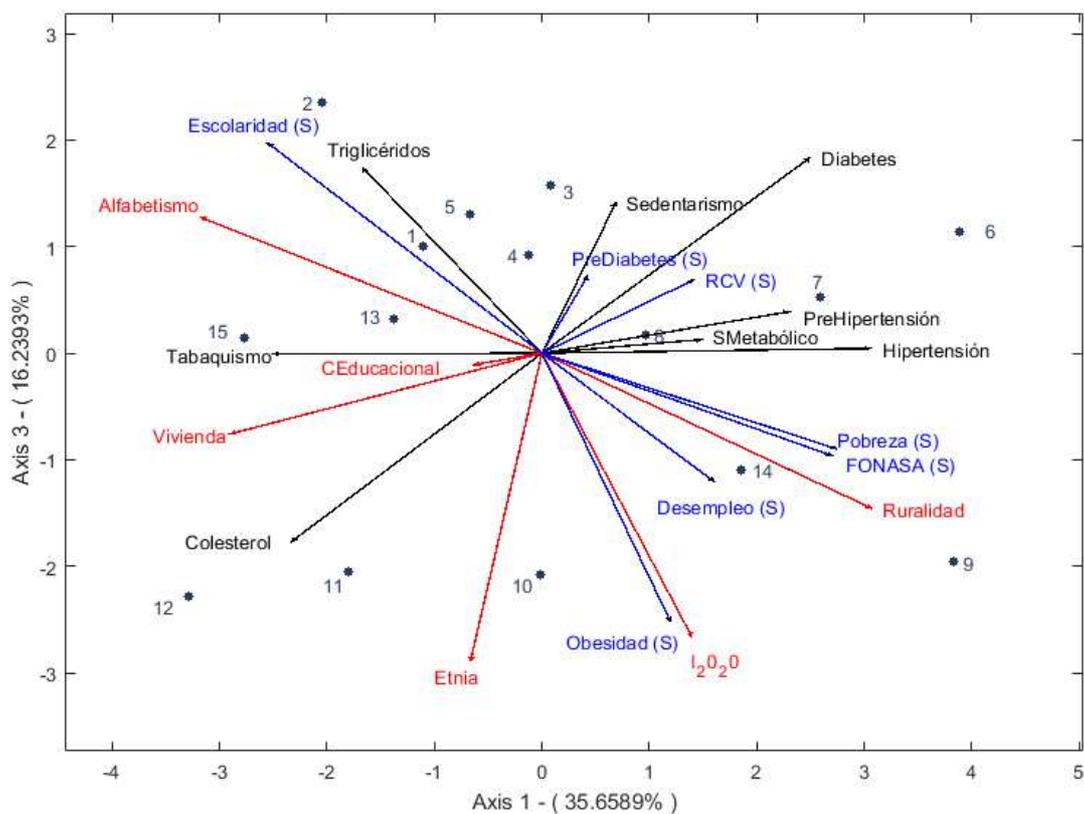


Figura 12. Plano 1-3. HJ BIPLLOT con Variables Suplementarias 2009

Para el año 2003 se realizó el mismo análisis que se decidió finalmente para el año 2009, esto es, se utilizaron las mismas variables activas y se representaron mediante un gráfico HJ BIPLLOT de componentes principales con variables suplementarias correspondientes a ese año. Los primeros dos factores explican un 59,68% de varianza total, mientras que al incluir el tercer factor se llega a 72,31%, por lo que nuevamente se mantienen 3 factores en el análisis.

Respecto a las contribuciones en el cuadro 8 se aprecia que en el plano 1-2 se debe interpretar colesterol, diabetes, tabaquismo, sedentarismo, ruralidad, etnia, alfabetismo, e índice 20/20, mientras que en el plano 1-3 se debe interpretar hipertensión, triglicéridos, síndrome metabólico, cobertura educacional, y calidad global de la vivienda.

Contribuciones			
Variable	Factor 1	Factor 2	Factor 3
Sospecha de hipertensión	302	5	113
Colesterol	4	501	1
Triglicéridos	537	6	356
Diabetes	483	1	0
Tabaquismo	575	159	41
Sedentarismo	268	110	0
Síndrome Metabólico	62	15	796
Cobertura Educativa	14	707	83
Calidad Global de la Vivienda	763	47	70
Ruralidad	827	82	19
Etnia	152	463	145
Alfabetismo	810	17	14
Índice 20/20	593	256	3

Cuadro 8. Contribuciones del factor al elemento columna análisis 2003.

En cuanto a la contribución de las filas, todas las regiones se ven representadas en el plano 1-2, excepto la quinta región que debe ser interpretada en el plano 1-3.

Contribuciones			
Región	Factor 1	Factor 2	Factor 3
I	712	30	17
II	525	104	60
III	306	30	79
IV	17	553	22
V	260	13	487
VI	92	597	137
VII	299	427	85
VIII	249	0	0
IX	474	240	194
X	747	2	9
XI	364	85	312
XII	376	314	16
RM	213	283	216

Cuadro 9. Contribuciones del factor al elemento fila análisis 2003.

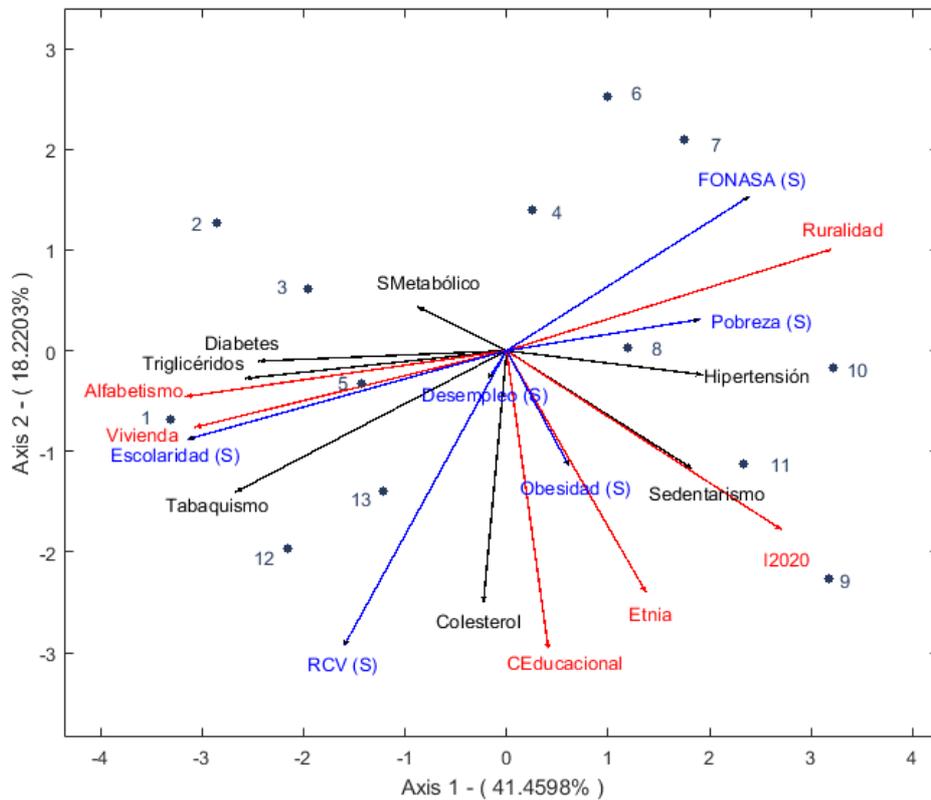


Figura 13. Plano 1-2. HJ BIPLLOT con Variables Suplementarias 2003

Para el año 2003, el plano 1-2 acumula un 59,67% (fig. 13) de variabilidad total de la representación obtenida.

Inmediatamente se puede apreciar que la estructura de comportamiento no es exactamente igual al año 2009, ya que no se interpretan los factores de la misma forma que el año anterior. Sin embargo, existen grupos de variables que se pueden interpretar en un mismo plano al igual que en 2009.

Al observar la figura 13 se aprecia que el tabaquismo tiene relación directa con el nivel de alfabetismo en la región. Tabaquismo también presenta una relación fuerte

e inversa con respecto a la ruralidad al igual que lo ocurrido en 2009 pero con una representación más baja en el plano 1-3. Se destaca que altos niveles de desigualdad de ingresos (índice 20/20) se relacionen de manera directa y fuerte con sedentarismo y en menor grado con colesterol muy similar a 2009. Además, la asociación del índice 20/20 con tabaquismo es casi nula (en 2009 fue inversa) y con diabetes se aprecia que la asociación es inversa. Por su parte al igual que en 2009, etnia tiene una asociación directa con colesterol, y esta vez en menor grado e inversamente con diabetes. Por otro lado, etnia no presenta relación importante con tabaquismo y si con sedentarismo. Este último factor de riesgo también se relaciona directamente con ruralidad, y de manera inversa con diabetes, tabaquismo y colesterol, así regiones con mayor cantidad de viviendas en zona rural presentan mayores niveles de riesgo mencionados anteriormente.

Respecto a las filas, nuevamente no tienen un parecido estrecho entre ellas. La IX región en 2003 se caracteriza con altos niveles de sedentarismo y desigualdad, además de etnia que extrañamente no se ve reflejado en el año 2009 a pesar de que en la actualidad se reconoce como una región con un porcentaje alto de personas pertenecientes a la etnia mapuche. La X y XI se caracterizan en el año 2003 por tener altos niveles de ruralidad, sedentarismo y desigualdad, lo mismo que ocurría en el año 2009 considerando que la región XV es un subconjunto de la X región. Por su parte las regiones II, III, IV, VI y VII no es posible caracterizarlas de manera apropiada ya que los vectores no se encuentran cerca de los puntos que

las representan. Finalmente, la XII y RM nuevamente presentan altos niveles de tabaquismo y niveles de colesterol importante.

Como era de esperarse la obesidad y el riesgo cardiovascular (ambas variables suplementarias) tienen relación directa con colesterol. Es importante mencionar que la obesidad, mostrada de manera pasiva, tiene relación con el índice de desigualdad 20/20.

El plano 1-3 acumula una varianza de 54,08% (figura 14). A primera vista se observa una relación inversa entre la calidad global de la vivienda y la hipertensión, ocurriendo lo mismo y de manera fuerte (ángulo de 180%) entre síndrome metabólico y cobertura educacional. El segundo punto llama la atención ya que es contrario a lo que ocurría en el 2009, donde si bien había una asociación esta era directa.

También es posible observar que la hipertensión se asocia a la cobertura educacional y síndrome metabólico con calidad de la vivienda de la misma manera que en el año 2009. Además, la pobreza es un factor importante en los niveles de hipertensión regional.

La V región se caracteriza solamente por determinantes sociales como la calidad global de la vivienda y la cobertura educacional.

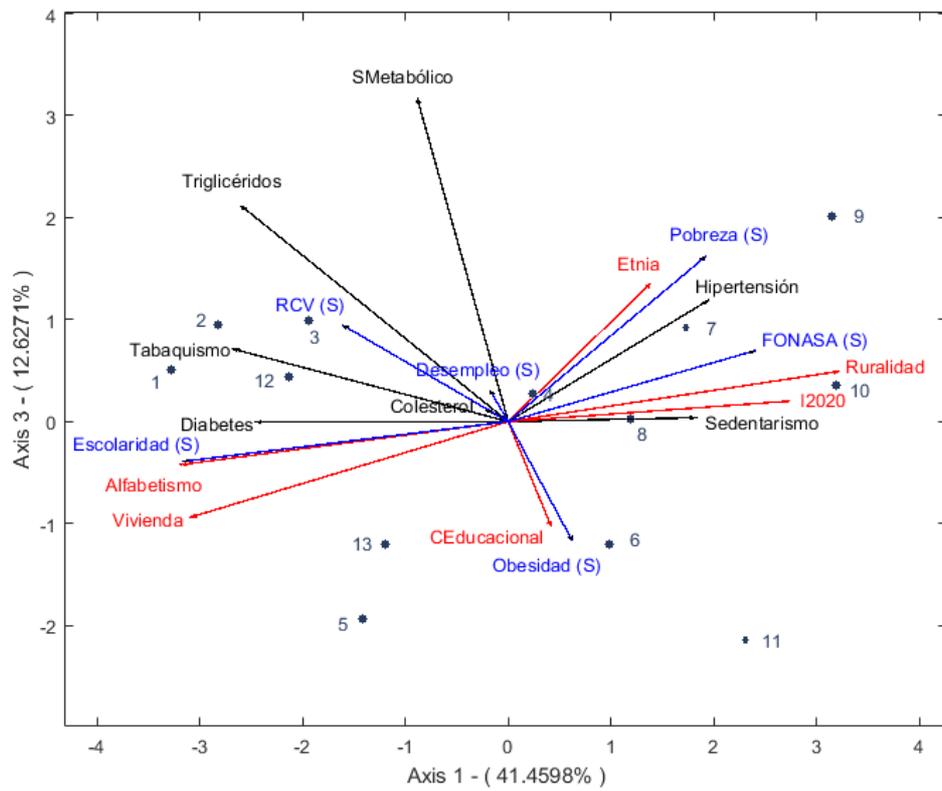


Figura 14. Plano 1-3. HJ BIPLLOT con Variables Suplementarias 2003

7 Capítulo VII: Discusión o Comentarios

En esta investigación se ha analizado la aplicación de un análisis multivariado representado en una gráfica denominada BIPLLOT apropiada con el fin de explicar e identificar la asociación existente entre factores de riesgo cardiovascular y determinantes sociales medidos para cada una de las regiones del país durante el año 2009, y su comparación descriptiva con estos indicadores en el año 2003.

La información recopilada mediante encuestas poblacionales; la ENS para variables de riesgo cardiovascular y CASEN en el caso de determinantes sociales resumen los indicadores seleccionados para cada una de las regiones del país. El análisis descriptivo univariado y el multivariado basado en análisis de componentes principales son de mucha utilidad cuando el foco se basa en la reducción de variables, por su parte los gráficos HJ-BIPLLOT resultan ser adecuados para explicar e identificar la asociación de factores de riesgo cardiovascular y determinantes sociales, además del comportamiento de las regiones acorde a estas variables, tanto por filas y columnas, consiguiendo una máxima calidad de representación.

Los determinantes sociales sí tienen relación con los factores de riesgo cardiovascular y las regiones del país se presentan un tanto heterogéneas con respecto a las variables. Los resultados obtenidos son consistentes con otras investigaciones, por ejemplo, estudios de prevalencia realizados en población indígena en Chile en zonas rurales en su mayoría han detectado prevalencias de diabetes de máximo 3,85% (47, 48), mientras que en zonas urbanas de 8,2% en

Mapuches y 6,9% en Aymaras (49) que puede ser explicado por el acceso a la mala alimentación en la ciudad. A pesar del alza sostenida en el tiempo, las prevalencias siempre son menores que las de la población general (ENS 2009-2010 es de 9,4%). Estos mismos estudios informan que los niveles de prevalencia de colesterol son más elevados que los de diabetes en las mismas poblaciones, y las prevalencias de colesterol son similares a la reportada a nivel nacional en la ENS 2009-2010 (50,51). Lo anterior condice a la asociación nula o inversa encontrada entre las variables diabetes y etnia, y a la asociación directa con colesterol presentada en los gráficos HJ BIPLLOT.

Determinantes sociales como la calidad global de la vivienda se relacionan con hipertensión de manera inversa y con tabaquismo de forma directa, cobertura educacional con síndrome metabólico (asociación directa) y pre hipertensión (asociación inversa) e índice 20/20 que se relaciona con síndrome metabólico, hipertensión y sedentarismo de forma directa.

Las variables suplementarias permiten incorporar interpretación a la gráfica BIPLLOT. En este caso, se aprecia la estrecha relación del índice 20/20 con la pobreza y la asociación directa que presentan con respecto al síndrome metabólico. Entre 2009 y 2003 existen diferencias y similitudes. De las diferencias destaca la asociación directa de tabaquismo con alfabetismo que en 2009 es totalmente contraria, además destaca que el índice 20/20 casi no se asocia con tabaquismo siendo que en 2009 si lo hace de manera inversa. Sin embargo, dado que la

representatividad de la base de datos en la ENS 2003 no es nivel región, este análisis es solo una referencia y no se pueden realizar grandes conclusiones al respecto. Respecto a las similitudes, se destaca la asociación existente entre etnia con respecto a diabetes y colesterol.

Una de las limitaciones más importantes en esta investigación es el número de observaciones con las que se cuenta, ya que también limita la cantidad de variables que se utilizan. Parece razonable instar a que se realicen investigaciones basadas en estos mismos análisis en un nivel más pequeño de observación, por ejemplo, provincia o comuna con el fin de caracterizar de mejor manera al país y poder trabajar con un número completo de variables. Lo anterior también permitiría utilizar otros tipos de análisis multivariado antes de la descomposición HJ BIPLLOT que logre no solo disminuir la dimensionalidad de las variables y presentarlas, sino que también incorporar análisis de regresiones, u otros métodos, además de cluster para las unidades de observación con el fin de indagar con mayor detalle respecto a los factores de riesgo cardiovascular y sus determinantes.

Otra oportunidad de investigación se refiere a la inclusión de otras patologías o problemas de salud que se aborden con este tipo análisis, dada la fácil y completa interpretación de los gráficos BIPLLOT es interesante explotar esta herramienta con el fin de presentar mayor información de manera sencilla y generar políticas relacionadas con salud.

8 Capítulo VIII: Conclusiones

La asociación encontrada entre determinantes sociales y factores de riesgo cardiovascular en Chile no se aleja de lo ya conocido respecto a las mismas.

Los determinantes sociales tales como; calidad global de la vivienda, desigualdad de ingresos (índice 20/20), cobertura educacional, etnia y ruralidad afectan los valores de los distintos factores de riesgo cardiovascular aumentando las prevalencias de síndrome metabólico, diabetes y colesterol, entre otros.

Por otro lado, a pesar de que los análisis descriptivos realizados primero de manera univariada y multivariada reflejaron cómo se comportan estas variables en las regiones de estudio, el uso de HJ BIPLLOT logra representar todo el análisis mediante una gráfica, llegando a conclusiones similares y aportando aún más información al problema estudiado, la cual se complementa aún más con la incorporación de variables suplementarias.

Si bien esta investigación no realiza grandes hallazgos respecto a riesgo cardiovascular y determinantes sociales, si incorpora una herramienta potente para descripción de fenómenos de este tipo, pudiendo explicar este problema de una forma clara y simple a un auditorio que no necesariamente se encuentra relacionado con la estadística.

9 Capítulo IX: Bibliografía

- (1) Lanas F, Serón P, Lanas A. Cardiovascular Disease in Latin America: The Growing Epidemic. *Progress in Cardiovascular Diseases*, 2014, 57(3): 262-267.
- (2) Reporte de Vigilancia de Enfermedades No Transmisibles (ENT), Santiago, Ministerio de Salud, 2011.
- (3) Atacan C, Alkan B, Sahin A. Examination of relation between nutrient components and fruits: Biplot approach. *International Journal of Food Sciences and Nutrition.*, 2009, 60(S1):181-189.
- (4) Cárdenas O, Galindo MP. Los métodos Biplot: evolución y aplicaciones. *Revista Venezolana de Análisis de Coyuntura*, 2007, 13(1): 279-303.
- (5) Otoo E, Asiedu R. The performance profile of *Dioscorea rotundata* cultivar Dorban genotypes in Ghana using GGE biplot analysis. *Journal of Food, Agriculture & Environment*, 2009, 7(1): 150-155.
- (6) Kruk C, Mazzeo N, Lacerot G, Reynolds S. Classification schemes for phytoplankton: a local validation of a functional approach to the analysis of species temporal replacement. *Journal of Plankton Research*, 2002, 24(9): 901-912.
- (7) Hinojosa L, Armesto J, Villagrán C. Are Chilean coastal forests pre-Pleistocene relicts? Evidence from foliar physiognomy, paleoclimate, and phytogeography. *Journal of Biogeography*, 2006, 33: 331-341.

- (8) Greenacre M. Biplots in Practice. 1° ed. España: Fundación BBVA, 2010.
- (9) Global status report on noncommunicable diseases 2010, Geneva, World Health Organization, 2011.
- (10) Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2001 to 2030. PLoS Med, 2006, 3(11): e442.
- (11) Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ, 2007, 335:136.
- (12) Padwal R, Straus SE, McAlister FA. Cardiovascular risk factors and their effects on the decision to treat hypertension: evidence based review. BMJ, 2001, 322:977.
- (13) Prevención de las enfermedades cardiovasculares: guía de bolsillo para la estimación y el manejo del riesgo cardiovascular, Ginebra, Organización Mundial de la Salud, 2008.
- (14) Pocock SJ, McCormack V, Gueyffier F, Boutitie F, Fagard RH, Boissel JP. A score for predicting risk of death from cardiovascular disease in adults with raised blood pressure, based on individual patient data from randomised controlled trials. BMJ, 2001,323:75.

- (15) Prevention of Cardiovascular Disease: Guidelines for assessment and management of cardiovascular risk, Washington, World Health Organization, 2007.
- (16) Global atlas on cardiovascular disease prevention and control. Geneva, World Health Organization, 2011.
- (17) Marti-Soler H, Gonseth S, Gubelmann C, Stringhini S, Bovet P, et al. Seasonal Variation of Overall and Cardiovascular Mortality: A Study in 19 Countries from Different Geographic Locations. PLoS ONE, 2014, 9(11): e113500.
- (18) Mejía-Lancheros C, Estruch R, Martínez-González MA, Salas-Salvado J, Castañero, et al. Impact of psychosocial factors on cardiovascular morbimortality: a prospective cohort study. BMC Cardiovascular Disorders, 2014, 14:135.
- (19) Darias S. Determinantes socioeconómicos y factores de riesgo cardiovascular: un estudio en canarias. Revista Atlántida, 2009, 1:93-113.
- (20) Krieger N, Waterman PD, Kosheleva A, Chen JT, Smith KW, et al. (2013) Racial Discrimination & Cardiovascular Disease Risk: My Body My Story Study of 1005 US-Born Black and White Community Health Center Participants (US). PLoS ONE 8(10): e77174.

- (21) Chobanian A, Bakris G, Black H et al. The Seventh Report of the Joint National Committee on Prevention, Detection and Treatment of High Blood Pressure, Hypertension, 2003, 42:1206-1252.
- (22) ALAD. Consenso de Prediabetes. Documento de Posición de la Asociación Latinoamericana de Diabetes, Revista de la Asociación Latinoamericana de Diabetes, 2009, 17:146-158.
- (23) II Encuesta de Calidad de Vida y Salud Chile 2006: Informe de resultados total nacional, Santiago, Ministerio de Salud, 2006.
- (24) Resultados I Encuesta de Salud: Chile 2003, Santiago, Ministerio de Salud, 2004.
- (25) Gabriel O, The biplot-graphic display of matrices with application to principal component analysis. *Biometrika*, 1971, 58: 453-467.
- (26) Torres-Salinas D, Robinson-García N, Jiménez-Contreras E, Herrera F, Delgado E. On the use of biplot analysis for multivariate bibliometric and scientific indicators. *Journal of the American Society for Information Science and Technology*, 2013, 64(7): 1468–1479.
- (27) Villardón J. Los métodos Biplot. Departamento de Estadística. Universidad de Salamanca, 2011.
- (28) Gower J, Gardner-Lubbe S, Le Roux N. *Understanding Biplots*. 1° ed. United Kingdom: Wiley; 2011.

- (29) Díaz-Faes A, González-Albo B, Galindo MP, Bordons M. HJ-Biplot como herramienta de inspección de datos bibliométricos. *Revista Española de Documentación Científica*, 2013, 36(1).
- (30) Cardenas O, Galindo MP. Biplot con información externa basado en modelos bilineales generalizados. Consejo de Desarrollo Científico y Humanístico. Universidad Central de Venezuela. Caracas, 2001.
- (31) Marcos Hidalgo M. HJ- Biplot Aumentado. Universidad de Salamanca, 2011.
- (32) Cárdenas O, Noguera C, Galindo MP, Vicente-Villardón J. Alternativa a la regresión con componentes principales basada en biplot de regresión. *Interciencia*, 2006, 31(3).
- (33) Fernández M, Galindo MP, Vicente-Villardón J, Martín A. Alternativa al análisis canónico de correspondencias basada en los métodos Biplot. *Mediterránea. Serie de Estudios Biológicos*, 1996, 15:63-71.
- (34) Graffelman J. Aluja Banet T. Optimal representation of supplementary variables in biplots from principal component analysis and correspondence analysis. *Biometrical Journal*, 2003, 45:485-509.
- (35) Graffelman J. Biplot with supplementary data. IX Conferencia Española de Biometría, 2003.
- (36) Galindo MP. Una alternativa de representación simultánea: HJ-Biplot. *Questíio*, 1986, 10(1):13-23.

- (37) Greenacre M, Blasius J. Multiple Correspondence Analysis and Related Methods. 1°Ed. U.S.: Chapman and Hall/CRC, 2006.
- (38) Peña D. Análisis de Datos Multivariantes. 1° ed. Madrid: McGraw-Hill; 2002.
- (39) Encuesta Nacional de la Salud. Chile 2009-2010, Santiago, Ministerio de Salud, 2010.
- (40) Marco Metodológico. Encuesta de Caracterización Socioeconómica Nacional. Casen 2003, Santiago, Ministerio de Desarrollo Social, 2005.
- (41) CASEN 2009. Informe Metodológico. Encuesta de Caracterización Social, Ministerio de Desarrollo Social, 2012.
- (42) Manual de uso de base de datos individual. Encuesta Nacional de la Salud 2009-2010, Santiago, Ministerio de Salud, 2012.
- (43) Criterios usados para la validación de base de datos de Encuesta Nacional de la Salud 2009-2010, Santiago, Ministerio de Salud, 2012.
- (44) Lê S, Josse J, Husson F. FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software, 2008, 25:1.
- (45) Vicente Villardón JL. MULTBILOT: A package for Multivariate Analysis using Biplots. Departamento de Estadística. Universidad de Salamanca. (<http://biplot.usal.es/ClassicalBiplot/index.html>), 2010.

- (46) Larenas G, Arias G, Espinoza O, Charles M, Landaeta O, Villanueva S et al. Prevalencia de diabetes mellitus en una comunidad Mapuche de la IX Región, Chile. *Rev Méd Chile*, 1985,113:1121-5.
- (47) Pérez-Bravo F, Carrasco E, Santos JL, Calvillán M, Larenas G, Albala C. Prevalence of type 2 diabetes and obesity in rural Mapuche population from Chile. *Nutrition*, 2001, 17: 236-8.
- (48) Carrasco E, Pérez F, Angel B, Cecilia A, Santos JL, Larenas G, Montalvo D. Prevalencia de diabetes tipo 2 y obesidad en dos poblaciones aborígenes de Chile en ambiente urbano. *Rev Méd Chile*, 2004, 132:1189-97.
- (49) Stockins B, Larenas G, Charles M, Standen D, Espinoza O, Illesca M et al. Niveles de lípidos y de presión arterial en población Mapuche de la región de la Araucanía, en Chile. *Rev Méd Chile*, 1998, 126: 1291-9.
- (50) Berríos X, Jadue L, Zenteno J, Ross MI, Rodríguez H. Prevalencia de factores de riesgo de enfermedades crónicas. Estudio en población general de la Región Metropolitana. 1986-1987. *Rev Méd Chile*, 1990, 118: 597-604.
- (51) Olaya J, Ippolito K, et al. Asociaciones entre la composición socioeconómica familiar y la generación urbana de residuos sólidos domiciliarios. *EIA*, 2014, 10(20): 127-137.

Anexo. Análisis de componentes principales para la reducción de variables

El Análisis de Componentes Principales (ACP) fue desarrollado por Hotteling en 1933 y busca reducir la dimensionalidad de una matriz de información que contiene variables cuantitativas, esto se refiere a la posibilidad de describir con precisión p variables mediante un conjunto reducido r ($r < p$). La utilidad de este método radica en que permite por un lado transformar las variables originales que generalmente están correlacionadas en nuevas variables incorreladas llamadas componentes principales con el fin de facilitar la interpretación de los datos; y, por otro lado, también es de utilidad para representar óptimamente en un espacio de dimensión pequeña la información (38). Existen tantas componentes principales como variables tenga el estudio, las que se obtienen calculando los valores propios λ_i de la matriz de varianzas y covarianzas, y sus vectores asociados, resolviendo (38):

$$|S - \lambda I| = 0 \quad (1)$$

$$(S - \lambda_i I)a_i = 0 \quad (2)$$

Si se llama a la matriz compuesta por columnas con valores de p componentes en n individuos, las nuevas variables están relacionadas con la matriz original como:

$$Z = AX \quad (3)$$

Donde $A'A = I$, el cálculo de los componentes principales corresponde a la transformación ortogonal de A a las variables X para obtener nuevas variables Z incorreladas (38).

Para decidir cuál es el aporte de cada factor se utiliza la varianza explicada. La varianza total explicada se define como la suma de los elementos de la diagonal principal de la matriz de correlaciones o traza de la matriz, luego cada componente aporta un porcentaje de ese total de varianza y ordenándolos de manera descendiente se escoge primero el factor que aporta mayor porcentaje, y así con cada uno de ellos (51). Así, si por ejemplo se desea utilizar un plano con dos componentes principales se espera que estos expliquen la mayor cantidad de varianza posible, alrededor de un 70% o más. De ser necesario se pueden incorporar más de 2 factores hasta llegar a un nivel acumulado de varianza explicada apropiado.

Como herramienta gráfica son de utilidad los denominados círculos de correlaciones, que, representados en el espacio generado por los dos primeros factores, sirven para analizar la varianza explicada de las variables utilizadas en el ACP, y de qué manera estas aportan a la estructura de los componentes (38, 51).

El uso de variables suplementaria cuando esta no se incluya en el ACP es importante para la interpretación, ya que una variable suplementaria o pasiva no afecta en el cálculo de los valores y vectores propios, pero reflejan una referencia de la asociación entre estas variables y los factores del análisis (51).