



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE TECNOLOGÍA

TEMA: TÉCNICAS DE BIG DATA A DATOS CLÍNICOS

AUTORES: Ing. CHÁVEZ VINUEZA, JORGE LUIS
Ing. ÑAÑAY PANCHO, JUAN MANUEL
Mat. MEDINA VÁSQUEZ, PAUL LEONARDO, Ph.D.

La motivación para realizar este Proyecto de Investigación es la obtención de conclusiones que posiblemente permitan mejorar el Estado de Salud de las personas, aprovechando la información contenida en la data, mediante la aplicación de herramientas y destrezas obtenidas en la MEMAT.

El presente trabajo de investigación desarrolla modelos matemáticos descriptivos y predictivos a partir de una data que contiene los resultados de exámenes clínicos de pacientes, en la misma se encuentra a individuos que mantienen una relación de dependencia con diversas empresas, es decir, la población de estudio es un segmento económicamente activo, con una relación laboral dependiente.

Pregunta de investigación

¿Es posible obtener información relevante, utilizando técnicas de Big Data, en la base de datos clínicos suministrados?

Para responderla, analizando los resultados de exámenes clínicos de pacientes, se utilizan las siguientes técnicas:

- ▶ Análisis descriptivo de las variables seleccionadas
- ▶ Análisis inferencial predictivo.

Big Data

Big Data es una combinación de tecnologías antiguas y nuevas que ayudan a obtener una perspectiva práctica. Es la capacidad de administrar un gran volumen de datos en diversos formatos, a la velocidad correcta y dentro de un marco de tiempo adecuado para permitir el análisis y la reacción en tiempo real

Las 5 Vs. del Big Data

Las características principales del Big Data son:

- ▶ **Volumen**, implica la cantidad de datos recolectados.
- ▶ **Variedad**, los datos se obtendrán en diversos formatos.
- ▶ **Velocidad**, la necesidad de procesar los datos en tiempos cortos.
- ▶ **Veracidad**, es el grado de confianza de la información obtenida de la data.
- ▶ **Valor**, que puede aportar la información a la actividad de la organización.

Preparación de la Data

- ▶ La data proporcionada es un historial **acumulado** entre los pacientes con el laboratorio clínico, así por ejemplo la información de un solo paciente consiste en una matriz de 500 filas y 60 columnas con mucha información redundante.
- ▶ La data proporcionada por el laboratorio se lo ordena en una nueva matriz. Una vez organizada la información se obtuvo una data con 23 336 observaciones y 104 variables de tipo cualitativo y cuantitativo.
- ▶ Se consideran las variables con información en el 15 por ciento o más de los pacientes, reduciéndose la data a 23 336 observaciones y 46 variables.
- ▶ Se considera únicamente la observación correspondiente a la última cita médica, reduciéndose la data a 18 657 observaciones y 46 variables.

Preparación de la Data (continuación)

- ▶ Se añaden 4 variables: dos que indican la actividad económica de la empresa donde labora el individuo (CIIU 40; CIIU 40 desagregado), una variable de identificación de las observaciones (ID), y la variable Edad. De esta manera se tiene una data de 18 657 observaciones y 50 variables.
- ▶ Las variables Colesterol LDL calculado, y Colesterol LDL cuantificado se fusionan en una sola variable llamada Colesterol LDL.
- ▶ Analizando las variables, una a una, se procede a detectar y eliminar datos inconsistentes.

Síndrome Metabólico Plus (SM+)

En este estudio nos centraremos en analizar el Síndrome Metabólico en el cual se considera las variables: Índice de Masa Corporal (IMC), Colesterol, Triglicéridos, Glucosa y Presión Arterial. Sin embargo, debido a la importancia de las variables Edad y los hábitos Sedentarismo y Tabaquismo, estas también serán consideradas, por tanto, a este conjunto de variables las llamaremos **Síndrome Metabólico Plus (SM+)**.

Variables consideradas

Variable	Unidades	Símbolo	Tipo
Índice de Masa Corporal	kg/m ²	IMC	cuantitativa
Colesterol	mg/dl	C	cuantitativa
Triglicéridos	mg/dl	Tri	cuantitativa
Glucosa	mg/dl	G	cuantitativa
Presión Sistólica	mmHg	Ps	cuantitativa
Presión Diastólica	mmHg	Pd	cuantitativa
Edad	años	E	cuantitativa
Sedentarismo (Hábito S)		S	cualitativa
Tabaquismo (Hábito T)		T	cualitativa
Sexo		Sx	cualitativa
Educación		Edu	cualitativa
Actividad Económica (CIU 40)		Ae	cualitativa
Identificación		ID	cualitativa

Actividades Económicas de la población

- ▶ La actividad económica donde se desenvuelve el paciente, definirá aspectos como su tipo de alimentación, nivel de actividad física diaria, actividad intelectual, nivel de estrés, etc., los mismos que influirán notablemente en su forma de vida y estado de salud.
- ▶ Dado que se está trabajando con una población económicamente activa, resulta de interés investigar cual es el comportamiento “de salud” en las diversas actividades económicas, pues en el caso de encontrar patrones particulares; esto, permitiría orientar recomendaciones de salud a cada una de ellas.

Actividades Económicas Totales de la Población

Actividades económicas en la Data	Población	Porcentaje
Agricultura, ganadería, silvicultura y pesca	35	0.28
Explotación de minas y canteras	3 329	26.93
Industrias manufactureras	336	2.72
Suministro de electricidad, gas, vapor y aire acondicionado	55	0.44
Distribución de agua; alcantarillado, gestión de desechos	44	0.36
Construcción	1598	12.93
Comercio al por mayor y al por menor	963	7.79
Transporte y almacenamiento	326	2.64
Información y comunicación	583	4.71
Actividades financieras y de seguros	2 199	17.79
Actividades inmobiliarias	5	0.04
Actividades profesionales, científicas y técnicas	1558	12.60
Actividades de servicios administrativos y de apoyo	708	5.73
Administración pública y defensa	15	0.12
Enseñanza	38	0.31
Actividades de atención de la salud humana	386	3.12
Otras actividades de servicios	161	1.30
Actividades de organizaciones y órganos extraterritoriales	24	0.19
Población total	12 363	100.00

Actividades Económicas Relevantes

Las principales actividades económicas, en porcentajes mayores al 10 % de la población total, que se encuentra en la data son:

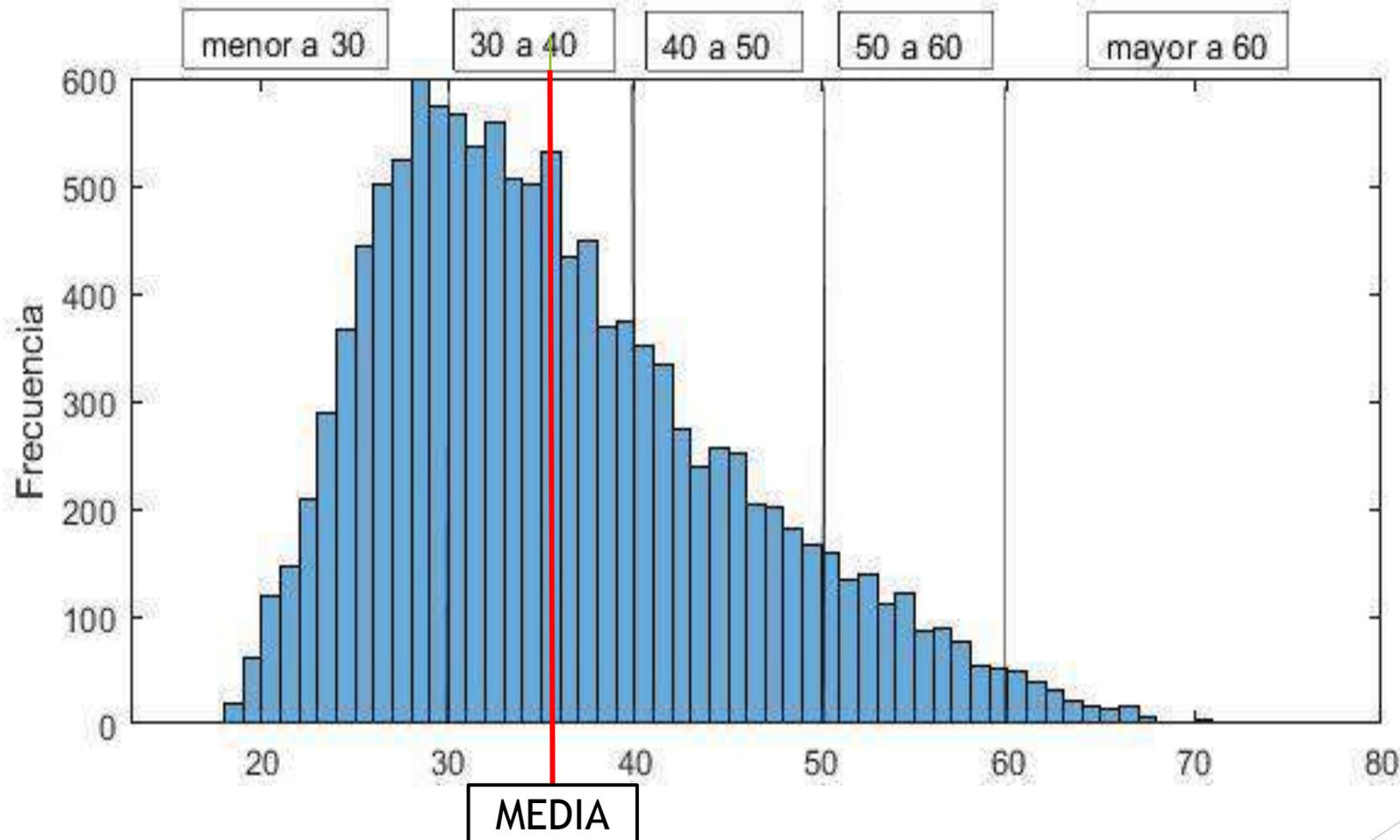
- ▶ Explotación de Minas,
- ▶ Actividades Financieras,
- ▶ Construcción,
- ▶ Actividades Profesionales y,
- ▶ Otras actividades



Análisis Estadístico Descriptivo

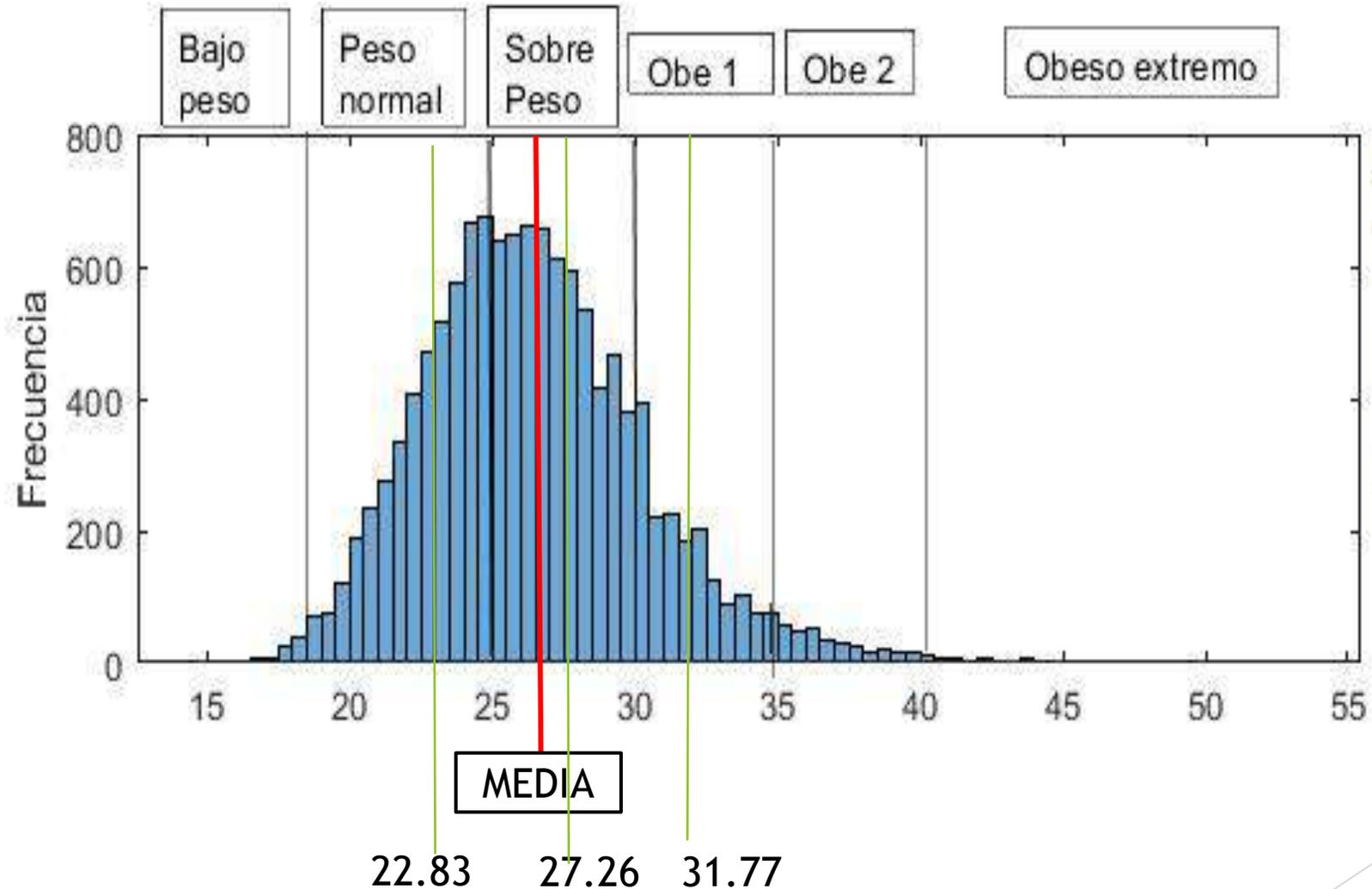


Estadísticos de la variable Edad, por categorías



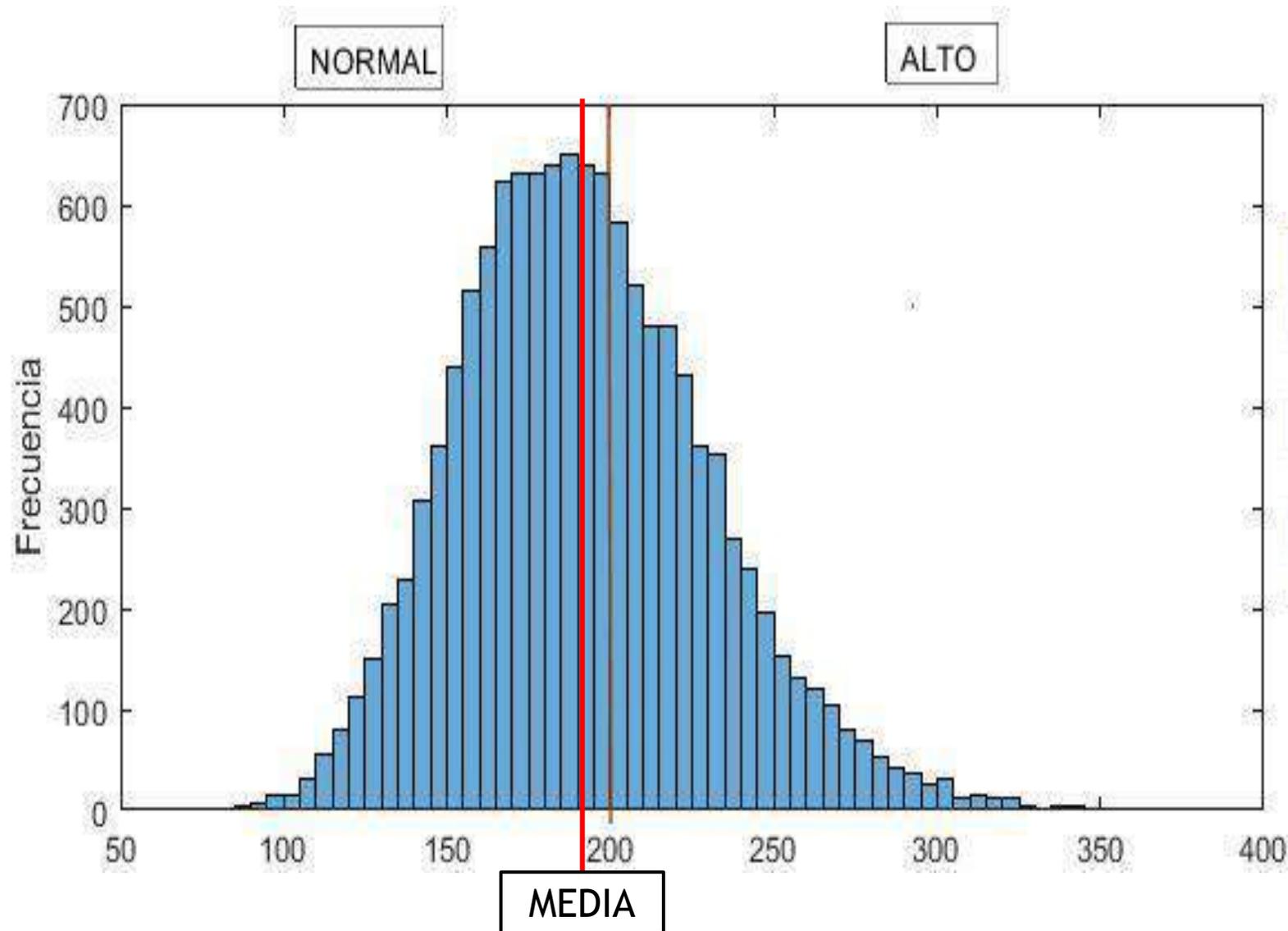
Estadístico	Valor	Unidades
media(μ)	35.49	años
mediana	34.00	años
varianza(σ^2)	91.87	años ²
d. estándar (σ)	9.58	años
asimetría	0.74	-
curtosis	3.10	-

Estadísticos de la variable IMC, por categorías.



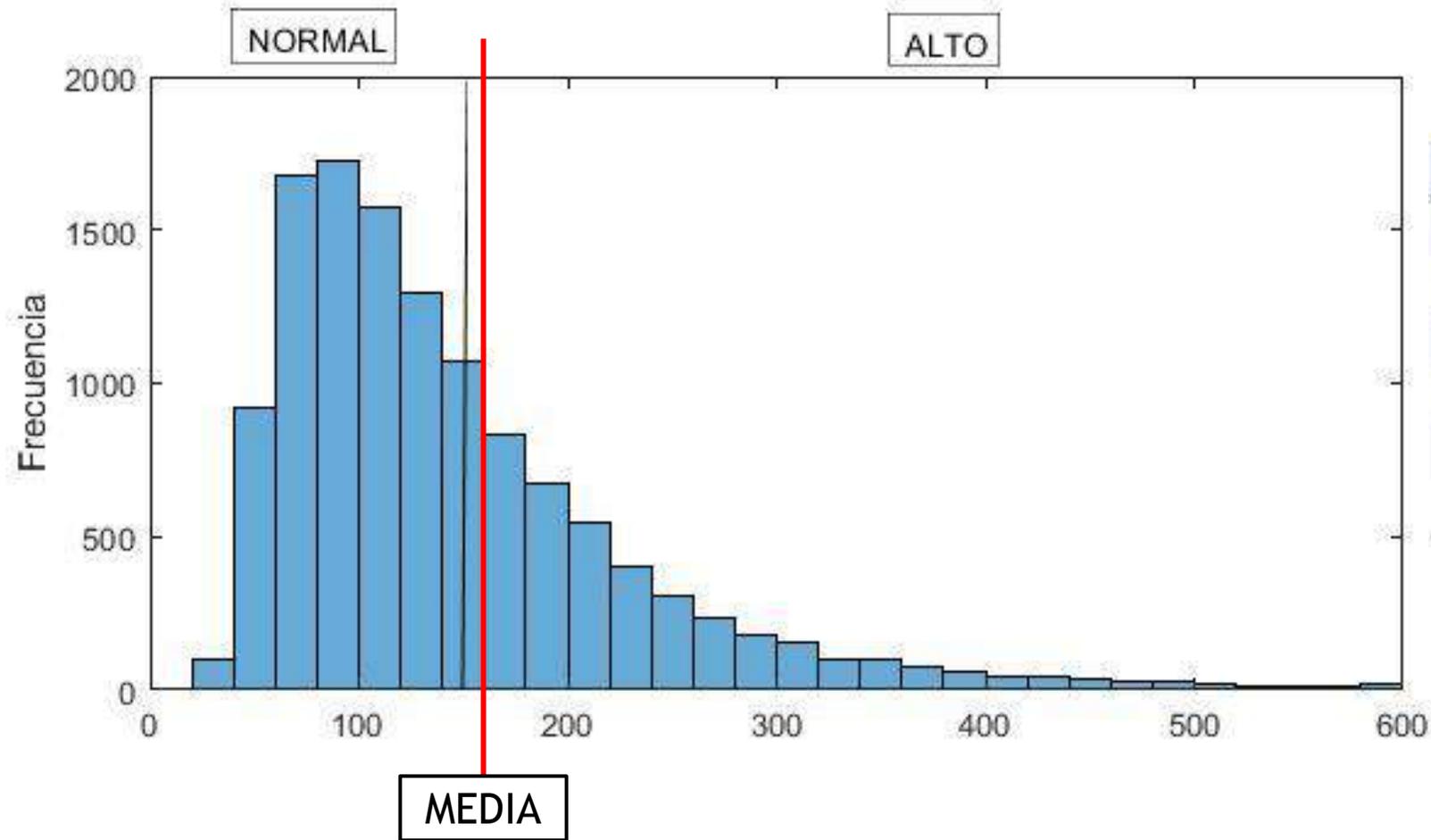
Estadístico	Valor	Unidades
media(μ)	26.48	kg/m^2
mediana	26.16	kg/m^2
varianza(σ^2)	15.74	kg^2/m^4
d. estándar (σ)	3.97	kg/m^2
asimetría	0.76	-
curtosis	4.59	-

Estadísticos de la variable Colesterol, por categorías.



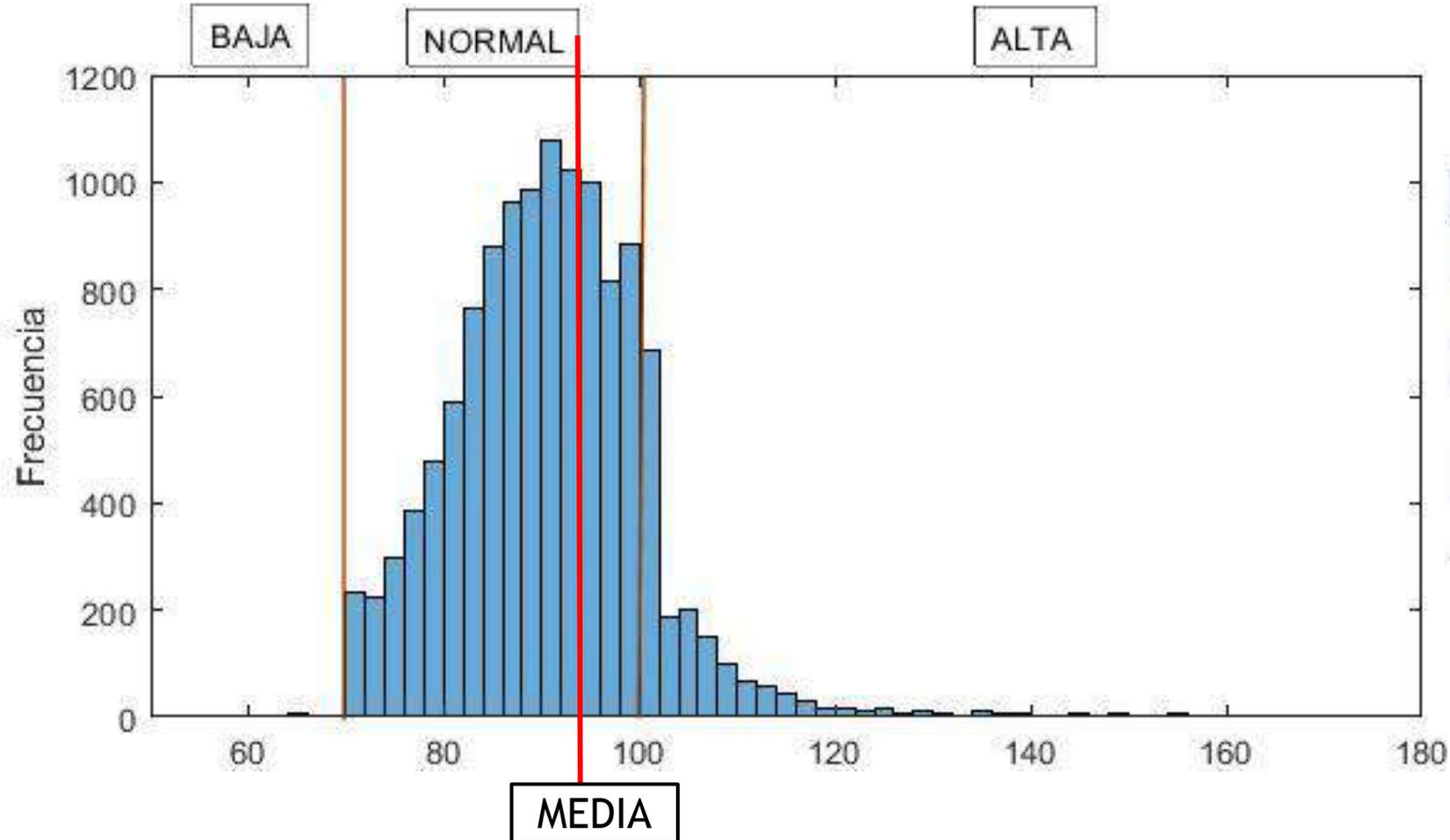
Estadístico	Valor	Unidades
media(μ)	191.78	mg/dl
mediana	189.00	mg/dl
varianza(σ^2)	1 558.00	mg ² /dl ²
d. estándar (σ)	39.47	mg/dl
asimetría	0.99	-
curtosis	12.57	-

Estadísticos de la variable Triglicéridos, por categorías.



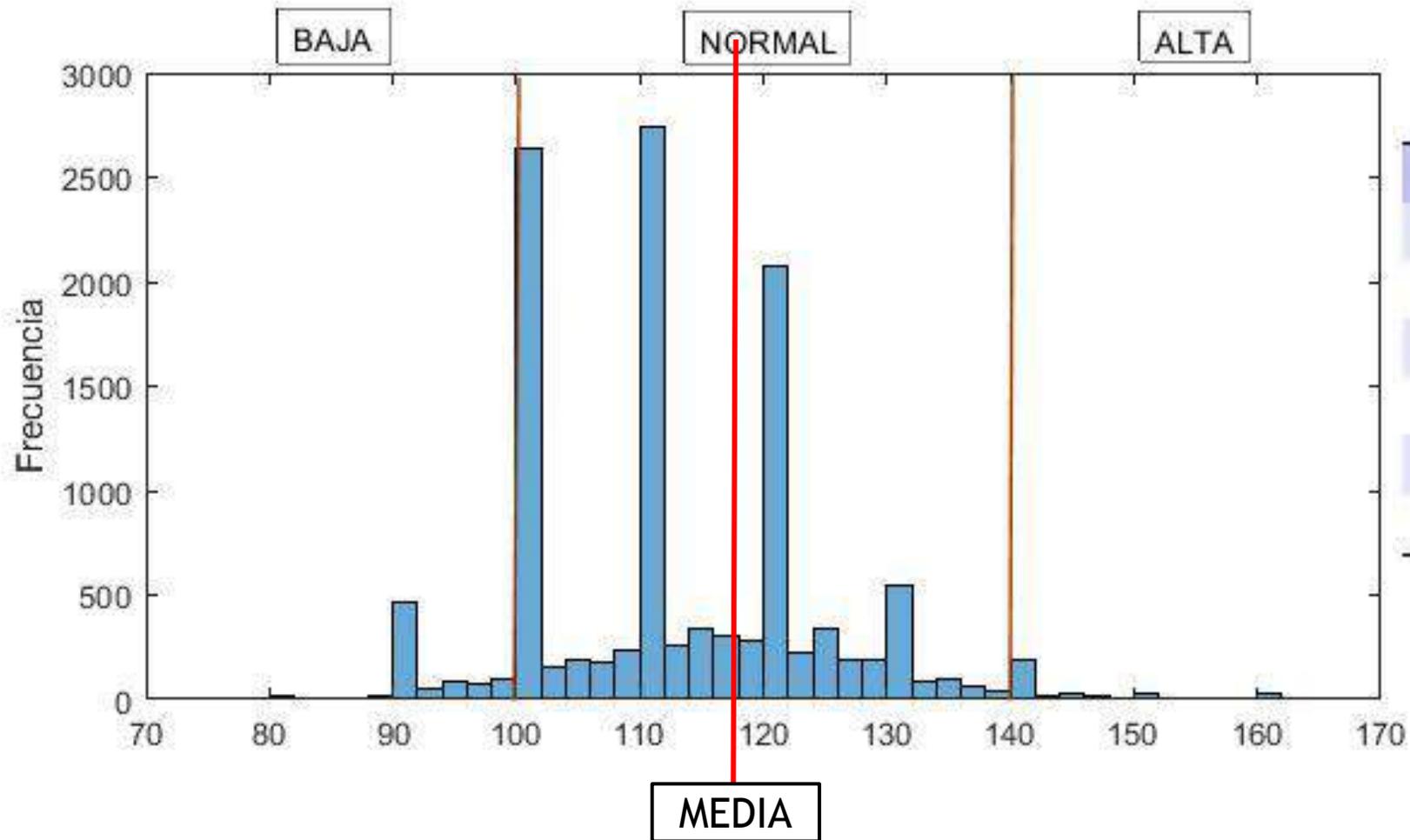
Estadístico	Valor	Unidades
media(μ)	149.21	mg/dl
mediana	122.00	mg/dl
varianza(σ^2)	11 743.00	mg ² /dl ²
d. estándar (σ)	108.36	mg/dl
asimetría	4.68	-
curtosis	51.90	-

Estadísticos de la variable Glucosa, por categorías.



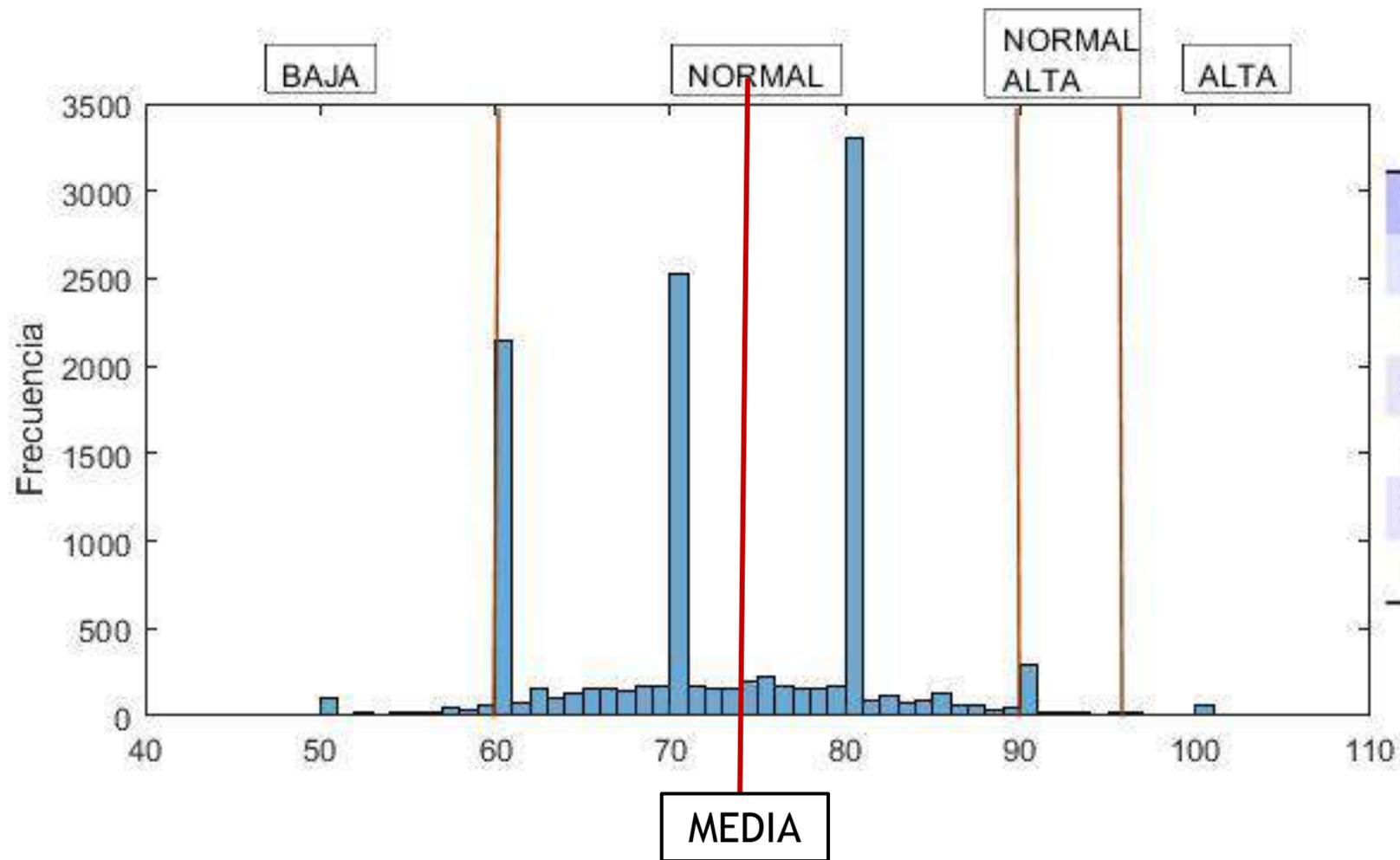
Estadístico	Valor	Unidades
media(μ)	91.03	mg/dl
mediana	90.00	mg/dl
varianza(σ^2)	241.31	mg ² /dl ²
d. estándar (σ)	15.53	mg/dl
asimetría	8.19	-
curtosis	136.36	-

Estadísticos de la variable Presión Sistólica



Estadístico	Valor	Unidades
media(μ)	112.29	mmHg
mediana	110.00	mmHg
varianza(σ^2)	150.66	mmHg ²
d. estándar (σ)	12.27	mmHg
asimetría	0.56	-
curtosis	3.91	-

Estadísticos de la variable Presión Diastólica



Estadístico	Valor	Unidades
media(μ)	72.40	mmHg
mediana	70.00	mmHg
varianza(σ^2)	88.05	mmHg ²
d. estándar (σ)	9.38	mmHg
asimetría	0.66	-
curtosis	3.88	-

Análisis Inferencial Predictivo



Análisis Inferencial Predictivo

El análisis de los datos se realiza de tres maneras:

- ▶ Determinando **Índices de Salud**. Aquí a las variables se las subdivide en categorías, según intervalos proporcionados por la literatura médica consultada.
- ▶ Hallando **factores de riesgo**. Aquí se transforma las variables cuantitativas en cualitativas dicotómicas.
- ▶ Encontrando las **similaridades** (distancias) entre los distintos individuos de un grupo de riesgo determinado.

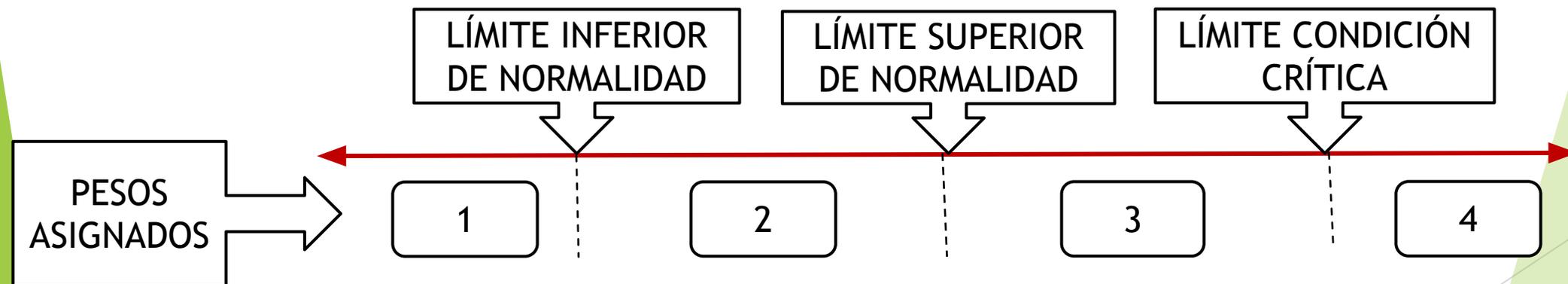
Índice de Salud

Para establecer el estado de salud de un paciente en función de los resultados exámenes clínicos de las variables: Colesterol, Triglicéridos, Glucosa y Presión Arterial, será importante construir un Índice de Salud.

El “**Índice de Salud**” de un paciente, es la suma de los pesos asignados a cada una de las distintas variables, de acuerdo al nivel de gravedad o riesgo.

Categorización de las variables en diversos niveles

- ▶ Las variables cuantitativas objeto de este estudio presentan diversos niveles de gravedad, por lo que se considera adecuado dividirlos en distintas categorías, de acuerdo al nivel de gravedad o riesgo. Los pesos asignados a las diversas categorías se indican en el siguiente gráfico.



Categorías de las variables: Colesterol, Triglicéridos, Glucosa, Presión Sistólica y Presión Distólica

Categoría	Símbolo	Rango	Valor
normal	C_n	<200	2
alto	C_a	≥ 200	3

Fuente: [Múnera and Escobar, 2007]

Categoría	Símbolo	Rango	Valor
normal	Tri_n	<150	2
alto	Tri_a	≥ 150	3

Fuente: [Calle, 2011]

Categoría	Símbolo	Rango	Valor
baja	G_b	<70	1
normal	G_n	[70, 100]	2
alta	G_a	≥ 100	3

Fuente: [Calle, 2011]

Categoría	Símbolo	Rango	Valor
baja	Ps_b	<100	1
normal	Ps_n	[100, 140]	2
alta	Ps_a	≥ 140	3

Fuente: [Rotaecche del Campo, 2002]

Categoría	Símbolo	Rango	Valor
baja	Pd_b	<60	1
normal	Pd_n	[60, 90]	2
normal alta	Pd_na	[90, 95]	3
alta	Pd_a	≥ 95	4

Fuente: [Rotaecche del Campo, 2002]

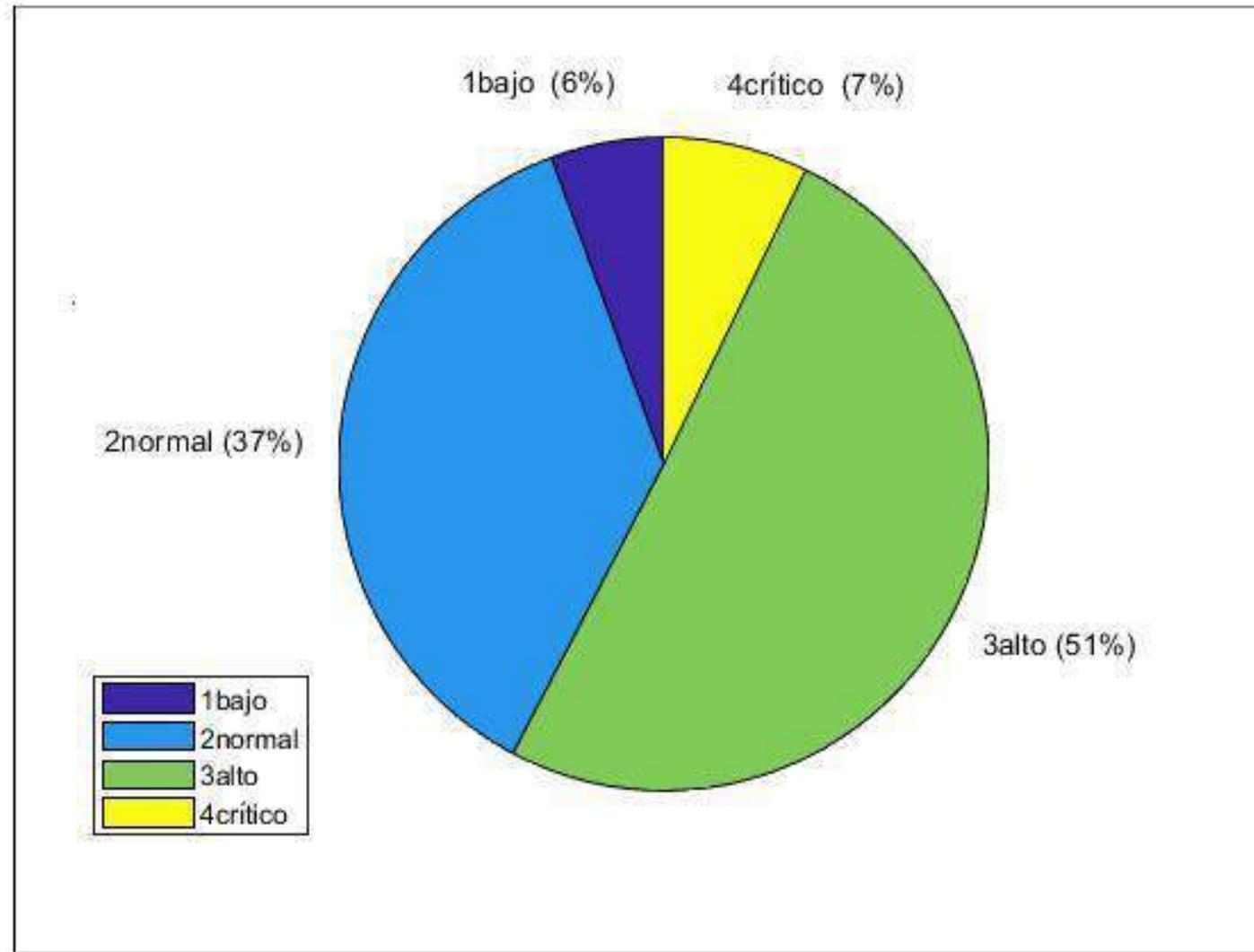
Estado del paciente de acuerdo al Índice de Salud

Estado de salud	Índice de salud
Bajo nivel de normalidad	8 a 9
Normal	10
Sobre nivel de normalidad	11 a 12
Crítico	13 a 16

Segmento de la data con variables valoradas, Índice de Salud y Estado de Salud

ID	Colesterol	Trigliceridos	Glucosa	Ps	Pd	Índice de Salud	Estado de Salud
1	3	3	3	2	2	13	Crítico
2	3	2	2	2	2	11	Sobre normal
3	2	3	3	2	2	12	Sobre normal
5	3	2	2	2	2	11	Sobre normal
6	2	2	2	2	2	10	Normal
10	3	2	3	2	2	12	Sobre normal
11	3	3	2	2	3	13	Crítico
12	2	2	2	2	2	10	Normal
13	3	3	2	2	2	12	Sobre normal
16	2	3	3	2	2	12	Sobre normal
17	3	2	2	2	2	11	Sobre normal
18	3	2	3	3	4	15	Crítico
19	3	3	3	2	2	13	Crítico
20	3	3	2	3	3	14	Crítico
23	3	3	2	2	2	12	Sobre normal
24	2	3	3	2	3	13	Crítico
26	3	2	3	2	2	12	Sobre normal
⋮							

Distribución de la población de acuerdo al Índice de Salud



Categorías de las variables IMC y Edad

Categorías para la variable Índice de Masa Corporal.

Categoría	Símbolo	Rango	Valor
bajo peso	IMC_b	<18.5	1
normal	IMC_n	[18.5, 25]	2
sobre peso	IMC_sp	[25, 30]	3
obeso 1	IMC_ob1	[30, 35]	4
obeso 2	IMC_ob2	[35, 40]	5
obeso extremo	IMC_ob3	≥40	6

Fuente: [OMS, 2018]

Categorías para la variable Edad.

Categoría	Símbolo	Rango	Valor
menores a 30 años	m30	<30	1
de 30 a 40 años	30_40	[30, 40]	2
de 40 a 50 años	40_50	[40, 50]	3
de 50 a 60 años	50_60	[50, 60]	4
mayores a 60 años	M60	≥60	5

Distribución de la población: variables múltiples categorías

Colesterol / Triglicéridos / Sedentarismo / IMC

	'1C_n'						'2C_a'						Total																														
	'1Tri_n'			'2Tri_a'			'1Tri_n'			'2Tri_a'																																	
Actividad e..	'0'						'1'																																				
'B - EXPLORACIÓN DE MINAS Y..	4	232	203	45	3	2	7	341	331	76	7	1	40	105	39	13	55	196	68	10	2	83	109	27	1	1	163	239	48	7	54	173	65	8	1	83	338	123	25	1	3.329		
'F - CONSTRUCCIÓN.'	40	30	10				6	367	267	43	8	2	3	19	7	3	1	48	124	45	3	3	13	15	6			92	116	31	6	1	7	17	6	2		99	168	44	4	2	1.598
'K - ACTIVIDADES FINANCIERAS..	9	198	114	30	2		9	414	280	62	20	4	36	54	24	4	36	101	41	7	2	2	62	48	13	1	2	96	112	33	8	36	59	14	3		66	131	57	5	3	2.199	
'M - ACTIVIDADES PROFESIONALES..	1	106	101	17	3	1	3	206	164	46	10	4	25	42	20	5	1	98	46	9	2	2	47	45	8	2	2	68	77	17	1	16	71	23	5	2	99	121	67	10	2	1.558	
'OTROS'	19	471	275	62	18	2	12	484	345	79	26	1	55	131	49	3	70	137	68	21	3	2	117	136	34	1	1	131	162	7	4	64	151	73	18	1	78	206	98	16	6	3.679	
	'1IMC_b'	'2IMC_n'	'3IMC_sp'	'4IMC_ob1'	'5IMC_ob2'	'6IMC_ob3'	'1IMC_b'	'2IMC_n'	'3IMC_sp'	'4IMC_ob1'	'5IMC_ob2'	'6IMC_ob3'	'1IMC_b'	'2IMC_n'	'3IMC_sp'	'4IMC_ob1'	'5IMC_ob2'	'6IMC_ob3'	'1IMC_b'	'2IMC_n'	'3IMC_sp'	'4IMC_ob1'	'5IMC_ob2'	'6IMC_ob3'	'2IMC_n'	'3IMC_sp'	'4IMC_ob1'	'5IMC_ob2'	'6IMC_ob3'	'2IMC_n'	'3IMC_sp'	'4IMC_ob1'	'5IMC_ob2'	'6IMC_ob3'	Total								

Explotación de Minas (%)

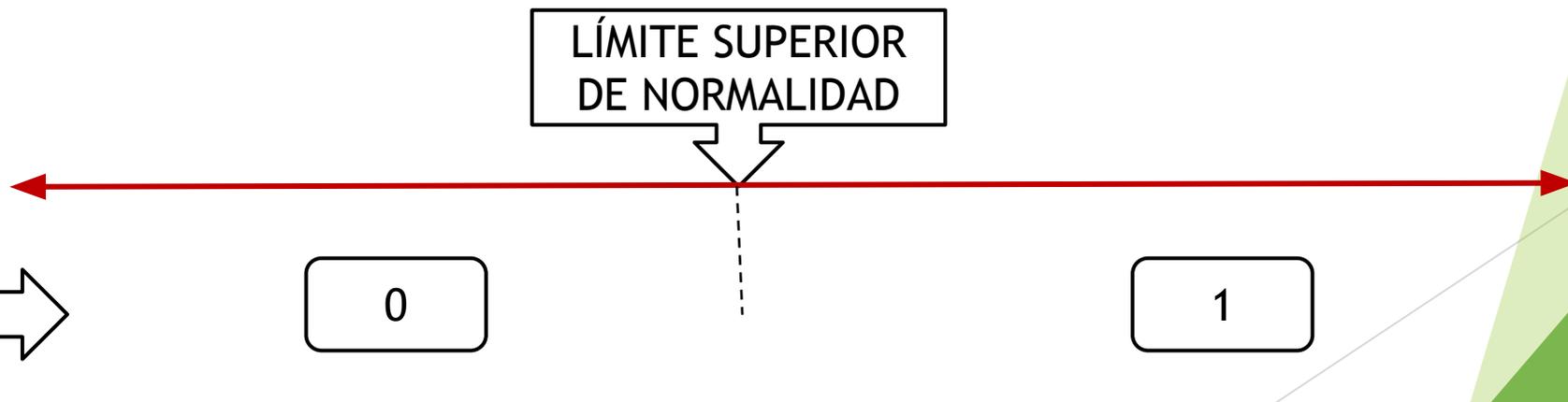


Actividades Financieras (%)



Factores de riesgo

Considerando el límite superior de normalidad de cada variable numérica, se construye la correspondiente variable dicotómica; de tal forma que 0, si no existe riesgo para la salud en esta variable y, 1 si esta variable representa un riesgo a la salud. A estas, sumamos las variables de Hábitos: Sedentarismo y Tabaquismo, formando de esta manera los nueve factores de riesgo que se analizan.



Límites de normalidad para las variables numéricas

Variable	Límite de normalidad	Unidades	Categoría 0	Categoría 1
E*	35	años	< 35	≥ 35
IMC	25	kg/m ²	< 25	≥ 25
G	100	mg/dl	< 100	≥ 100
C	200	mg/dl	< 200	≥ 200
Tri	150	mg/dl	< 150	≥ 150
Ps	140	mmHg	< 140	≥ 140
Pd	90	mmHg	< 90	≥ 90

Fuente: [OMS, 2018], [Múnera and Escobar, 2007], [Calle, 2011], [Rotaecbe del Campo, 2002].

Grupos de riesgo

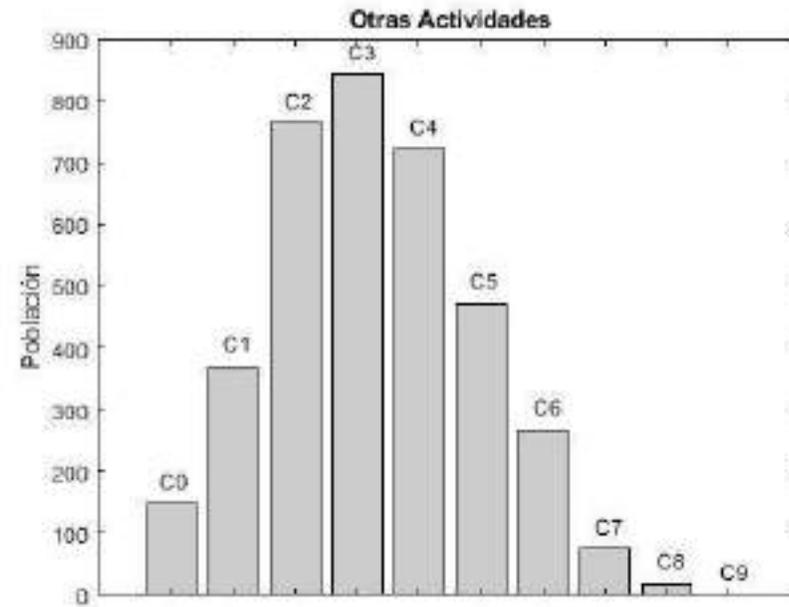
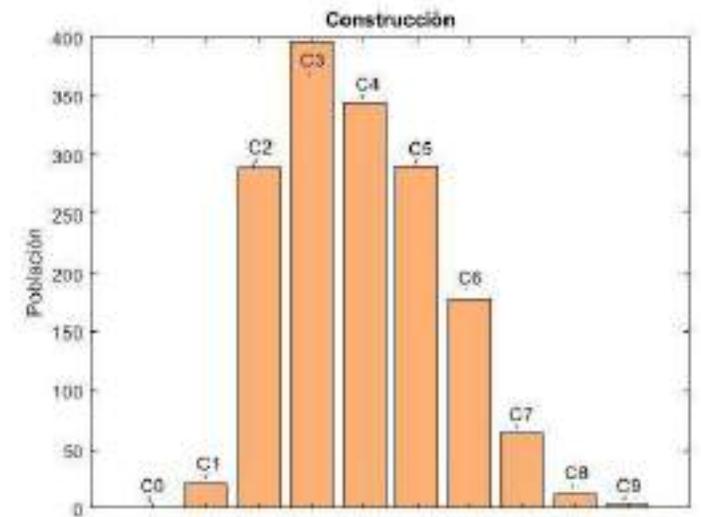
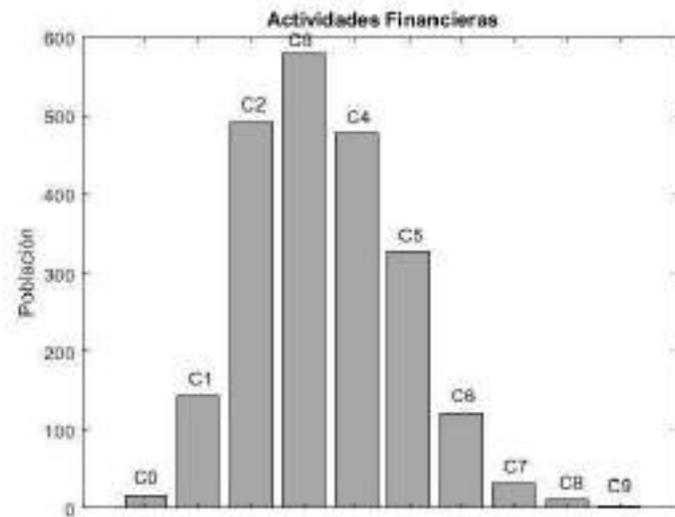
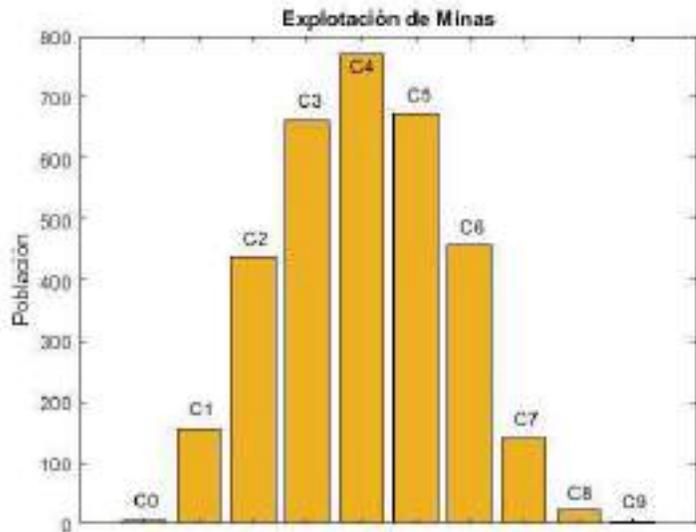
Los grupos de riesgo denominados C_i ; $i = 0, 1, \dots, 9$; corresponden a grupos poblacionales que poseen i factores de riesgo, los mismos que pueden ser cualquiera de las nueve variables analizadas.

Distribución de la población por Actividad Económica y número de Factores de Riesgo

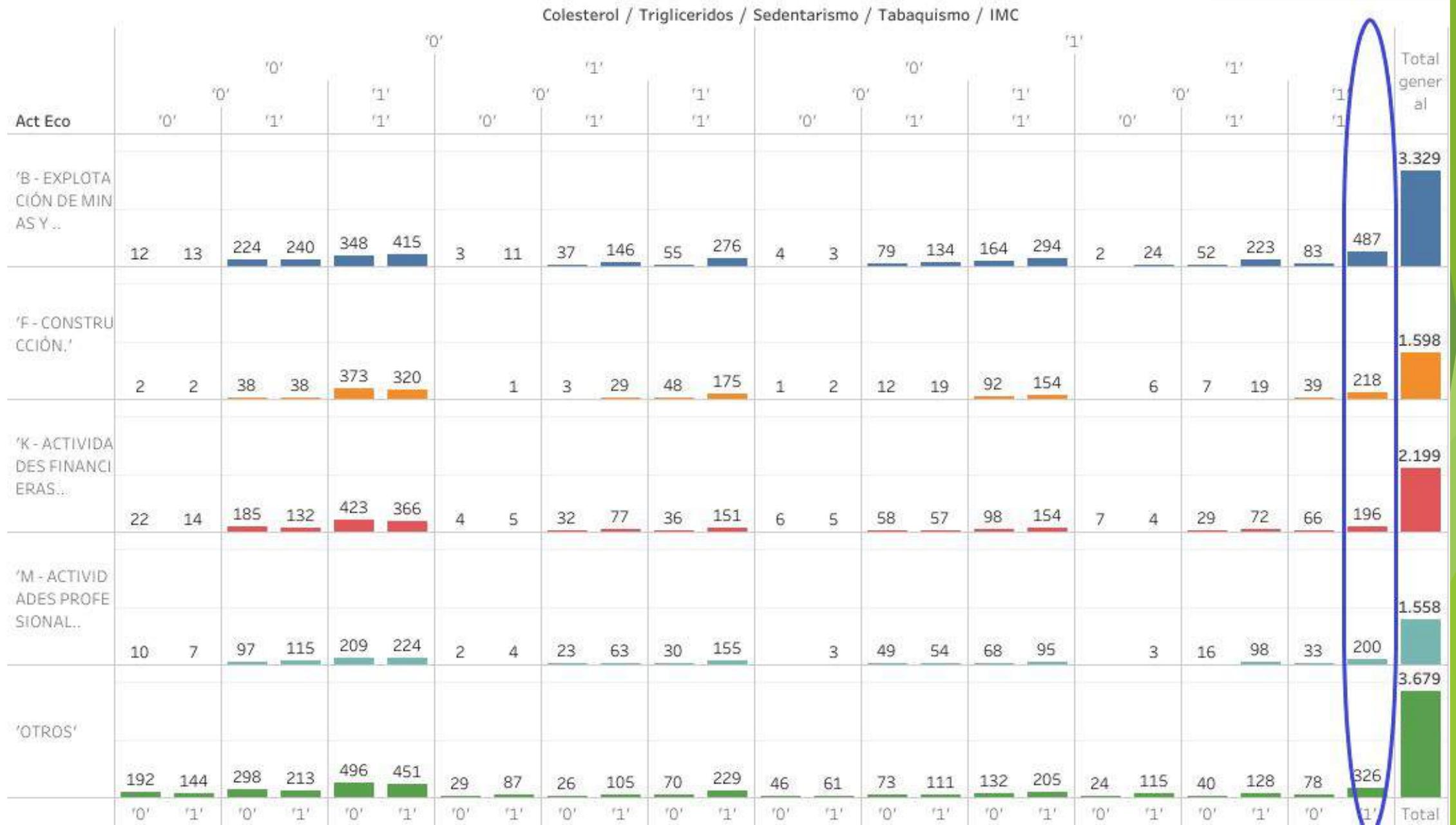
Grupos de riesgo	Explotación de Minas		Actividades Financieras		Construcción		Actividades Profesionales		Otras actividades		Total
	población	%	población	%	población	%	población	%	población	%	
C_0	5	0.15	16	0.72	1	0.06	9	0.57	149	4.05	180
C_1	155	4.65	142	6.45	22	1.37	76	4.87	369	10.03	764
C_2	437	13.12	490	22.28	289	18.08	267	17.13	764	20.77	2247
C_3	661	19.85	579	26.33	396	24.78	353	22.65	844	22.94	2833
C_4	771	23.16	477	21.69	344	21.52	346	22.20	722	19.62	2660
C_5	674	20.24	326	14.82	290	18.14	292	18.74	470	12.78	2052
C_6	458	13.75	122	5.54	178	11.13	143	9.17	267	7.26	1168
C_7	142	4.26	32	1.45	63	3.94	55	3.53	74	2.01	366
C_8	23	0.69	12	0.54	12	0.75	15	0.96	18	0.49	80
C_9	3	0.09	3	0.13	3	0.18	2	0.12	2	0.05	13
Total	3 329	100.00	2 199	100.00	1 598	100.00	1 558	100.00	3 679	100.00	12 363

C_i = grupos con i factores de riesgo.

Actividades Económicas por número de Factores de Riesgo

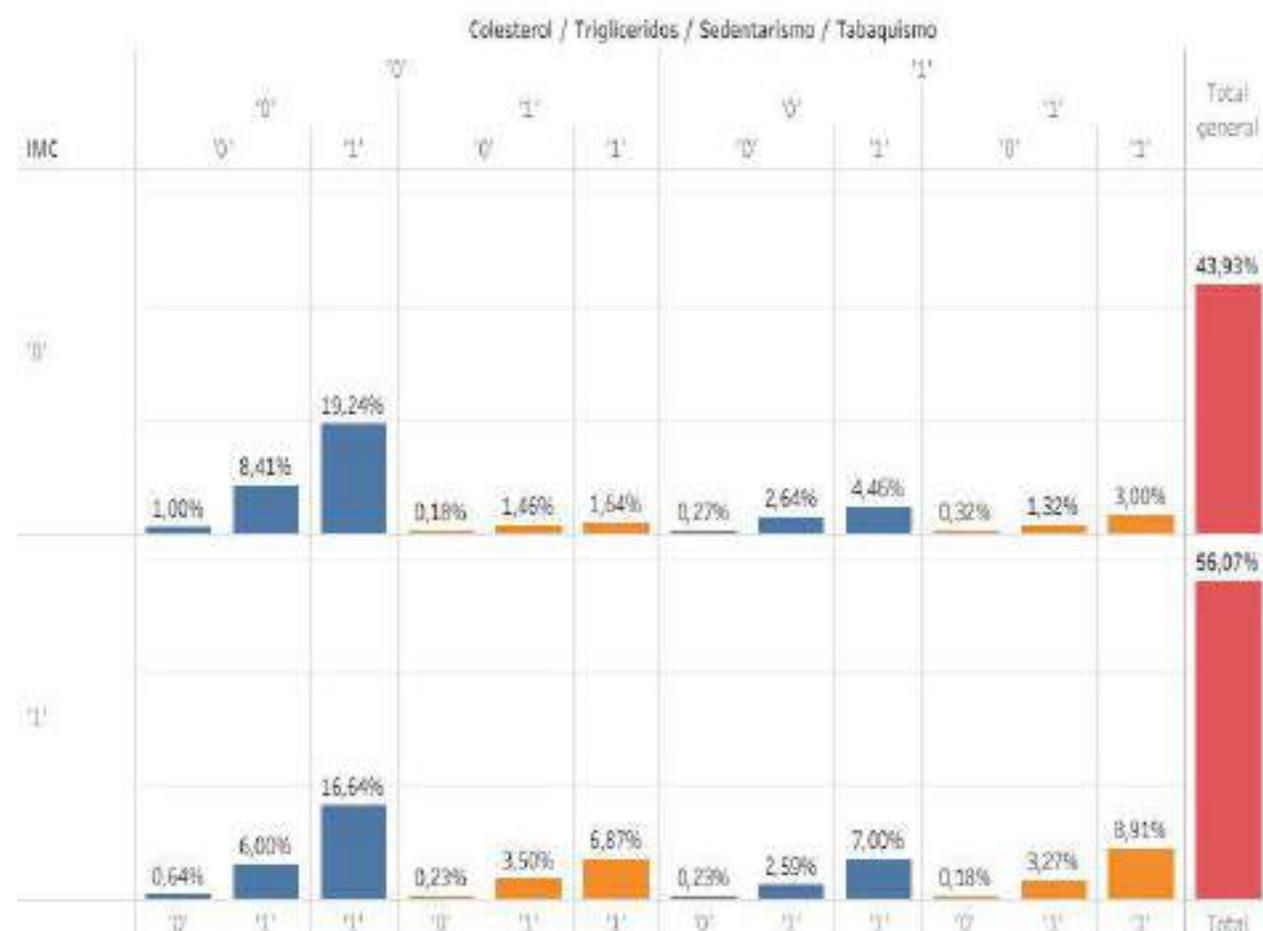
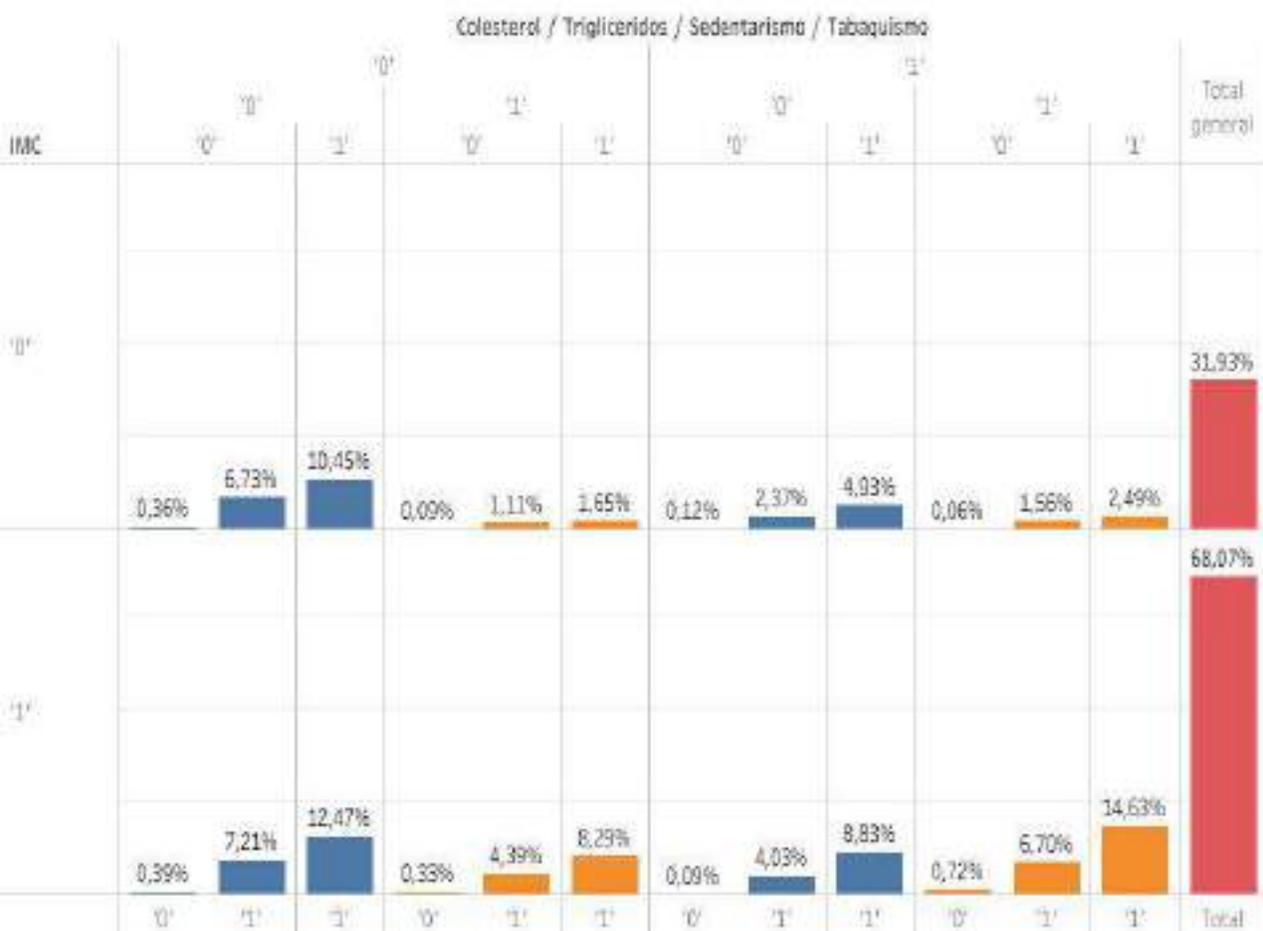


Distribución de la población: variables dicotómicas



Explotación de Minas (%)

Actividades Financieras (%)



Escalado Multidimensional

Se aplica esta técnica a dos grupos de actividades económicas, Explotación de Minas y Actividades Financieras; ambos con tres factores de riesgo.

Variables consideradas para el Escalado Multidimensional

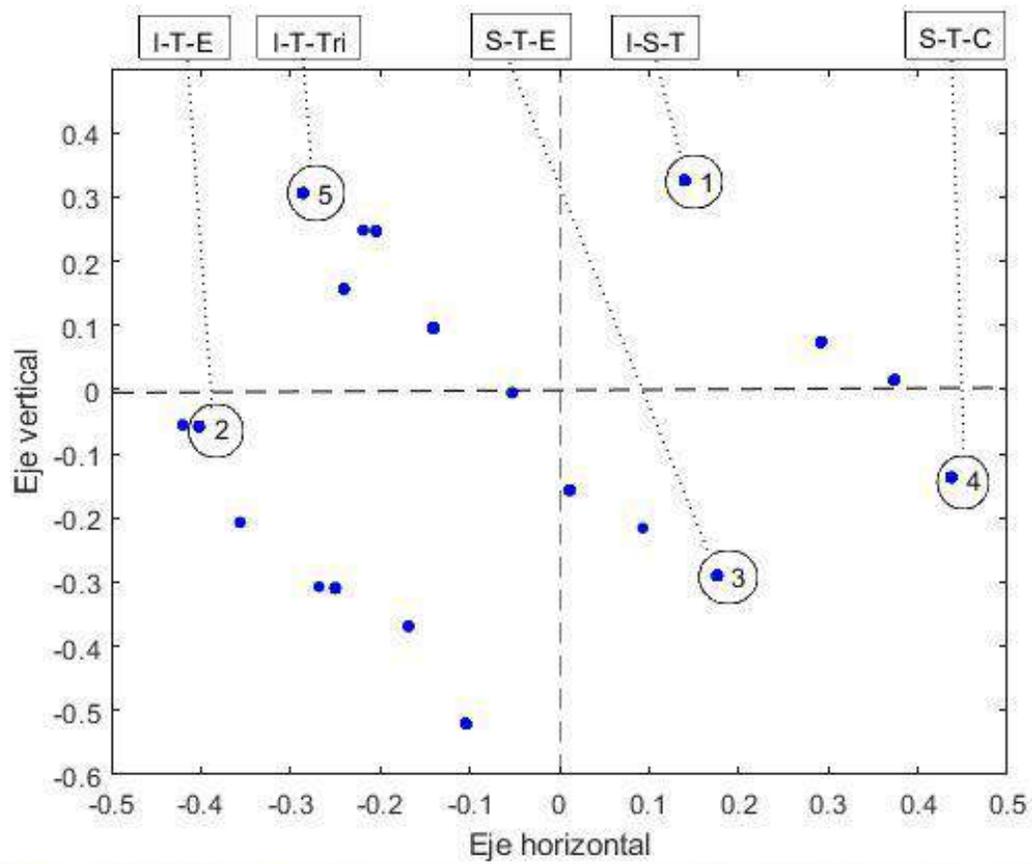
Número	Factor de Riesgo	Simbología	Tipo	Categoría
1	Índice de Masa Corporal	IMC	Mecánicos	F1
2	Presión Sistólica	Ps	Mecánicos	F1
3	Presión Diastólica	Pd	Mecánicos	F1
4	Colesterol	C	Químicos	F2
5	Triglicéridos	Tri	Químicos	F2
6	Glucosa	G	Químicos	F2
7	Sedentarismo	S	Hábitos	F3
8	Tabaquismo	T	Hábitos	F3
9	Edad	E	Edad	F4

Aplicación del Escalado multidimensional

- ▶ En los grupos escogidos, se aplica el coeficiente de similaridad, y se obtiene matrices de similaridades.
- ▶ A partir de las matrices de similaridades, se obtiene las matrices de distancias.
- ▶ En las matrices de distancias aplicando el proceso de escalado multidimensional, se representa cada una de las observaciones del grupo mediante dos nuevas variables ortogonales, pudiendo visualizar a cada una de las observaciones como puntos en un plano.
- ▶ Cada una de estas nuevas variables ortogonales son en realidad una combinación lineal de las variables originales sin que ninguna de estas tenga una única preponderancia sobre el resto, por lo que en este estudio las llamaremos como Eje horizontal y Eje vertical.

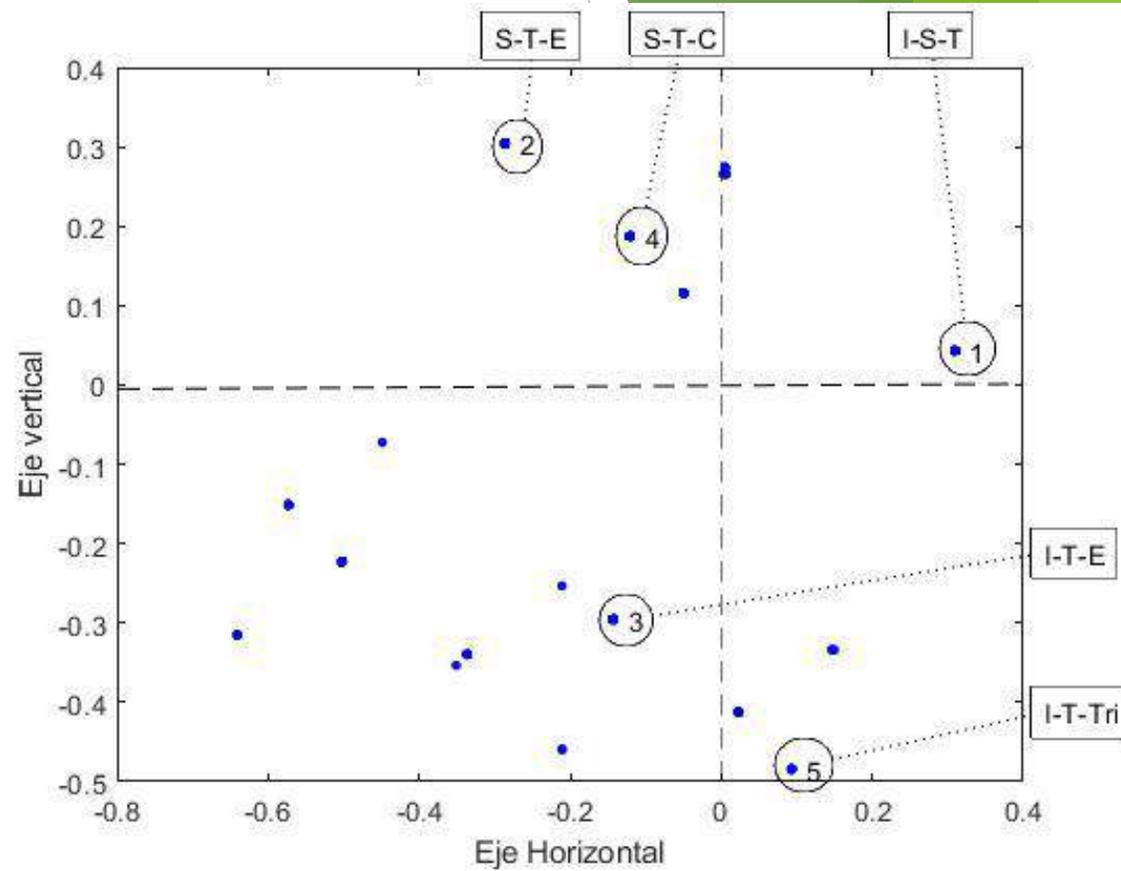
Similaridades: Explotación de Minas vs. Actividades Financieras

Explotación de Minas



Observación	Factor 1	Factor 2	Factor 3	Porcentaje
1	IMC	Sedentarismo	Tabaquismo	24
2	IMC	Tabaquismo	Edad	17
3	Sedentarismo	Tabaquismo	Edad	15
4	Colesterol	Sedentarismo	Tabaquismo	10
5	IMC	Triglicéridos	Tabaquismo	5

Actividades Financieras

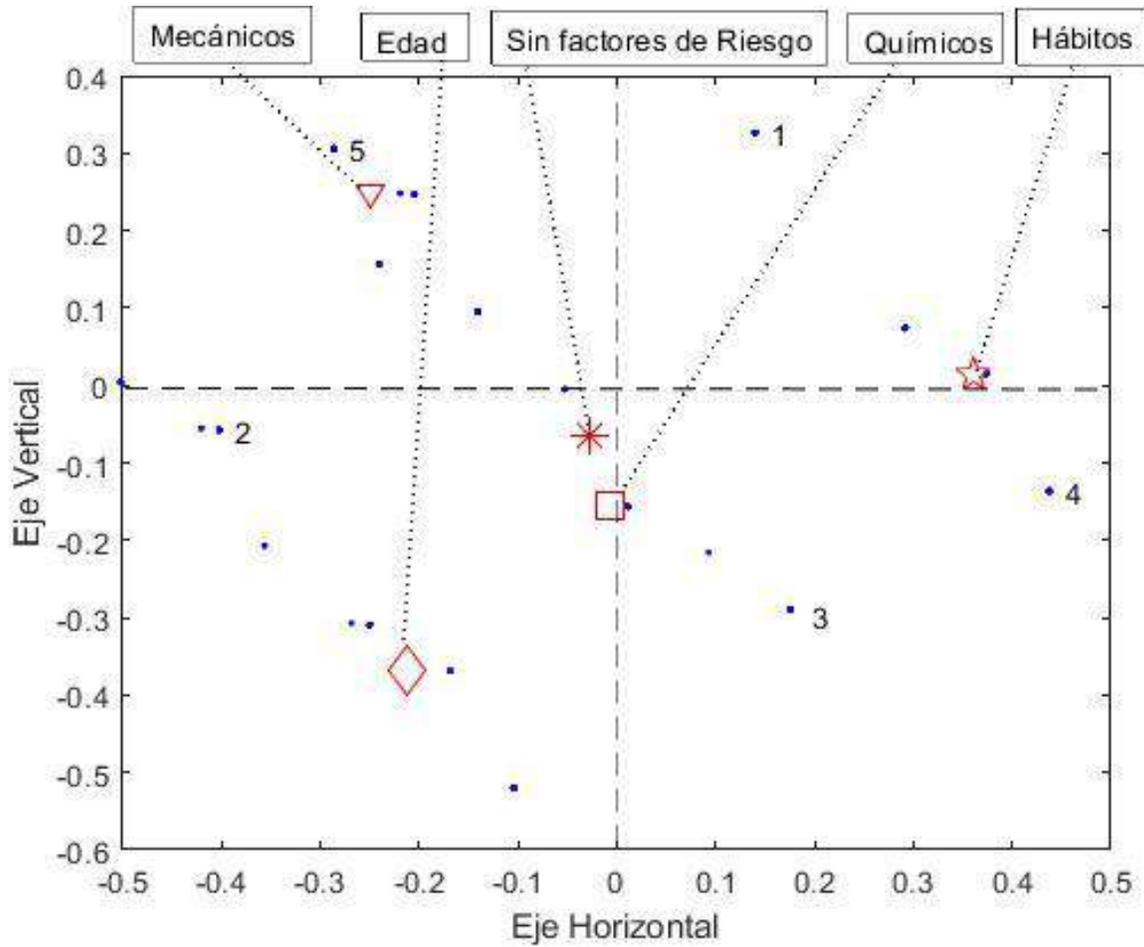


Observación	Factor 1	Factor 2	Factor 3	Porcentaje
1	IMC	Sedentarismo	Tabaquismo	39
2	Sedentarismo	Tabaquismo	Edad	17
3	IMC	Tabaquismo	Edad	8
4	Colesterol	Sedentarismo	Tabaquismo	8
5	IMC	Triglicéridos	Tabaquismo	5

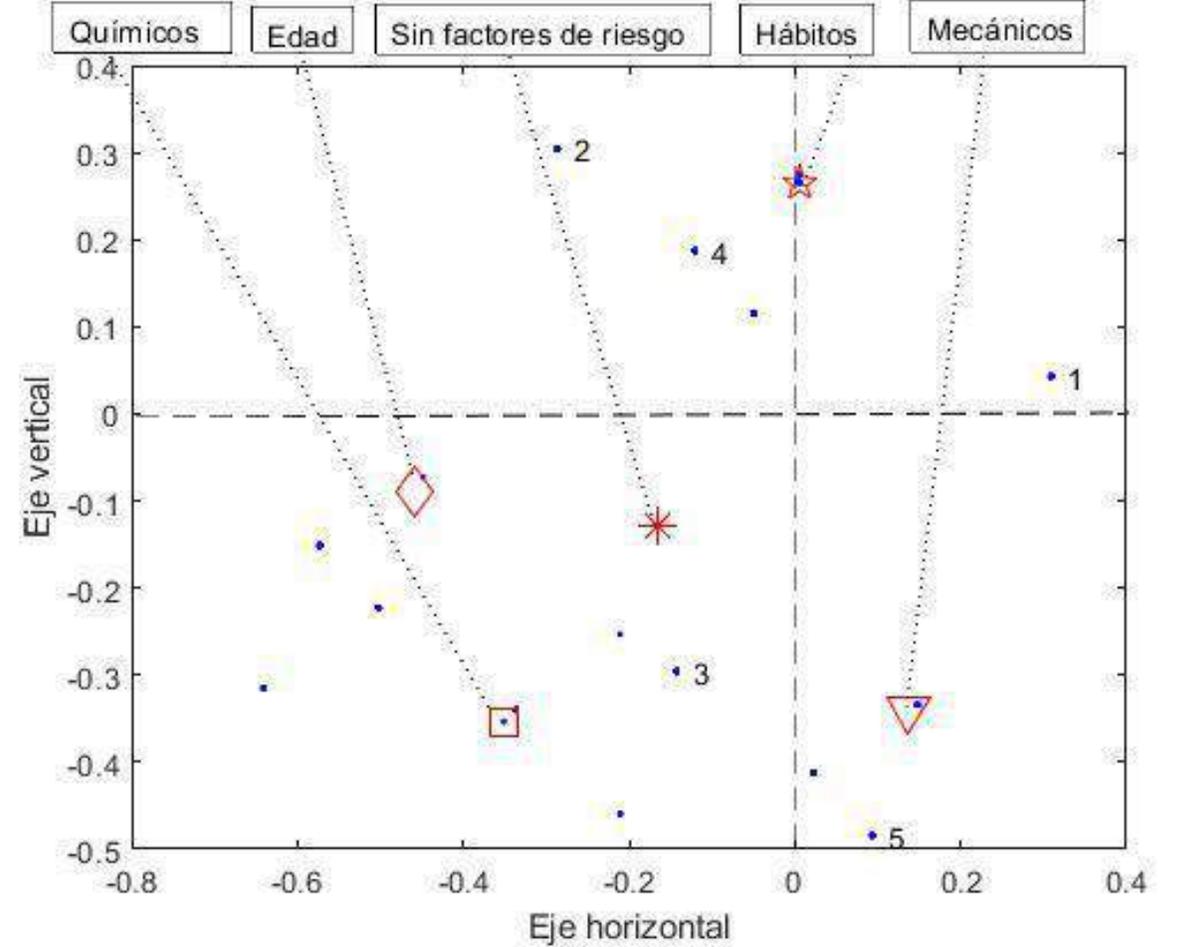
Comentario

Los gráficos de similitudes de los grupos Explotación de Minas y Actividades Financieras presentan ciertas diferencias, las que se visualizan de mejor modo con la información de las tablas correspondientes a los factores de riesgo de los subgrupos más representativos y el porcentaje de estos respecto a la población total.

Representación de factores: Mecánicos, Químicos, Hábitos y Edad

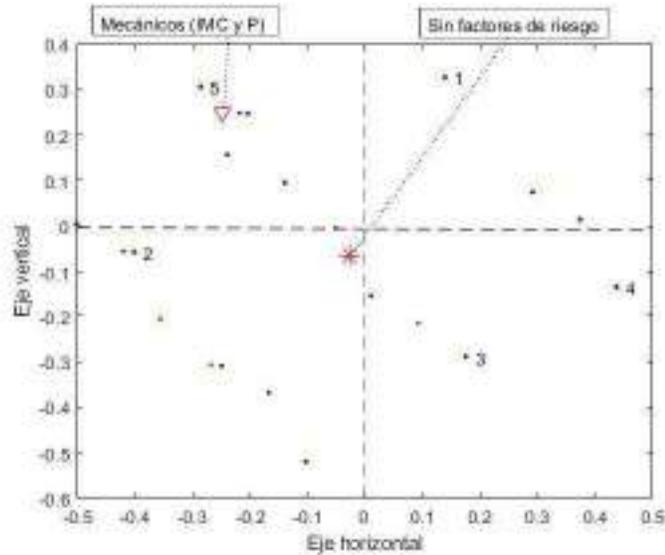


Minas

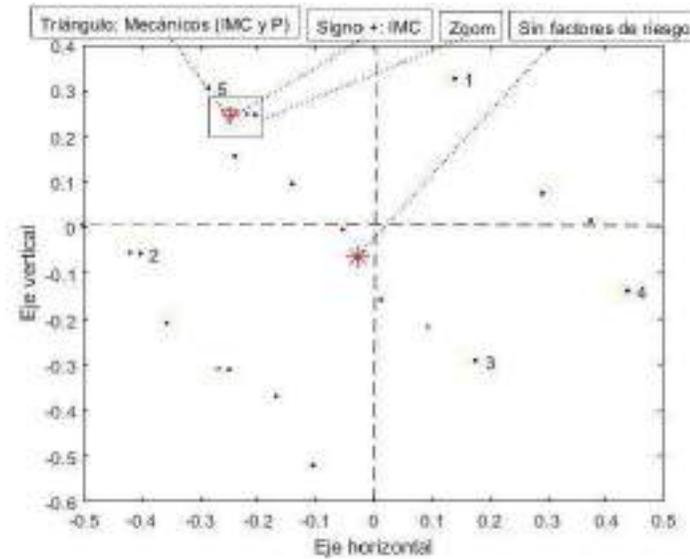


Finanzas

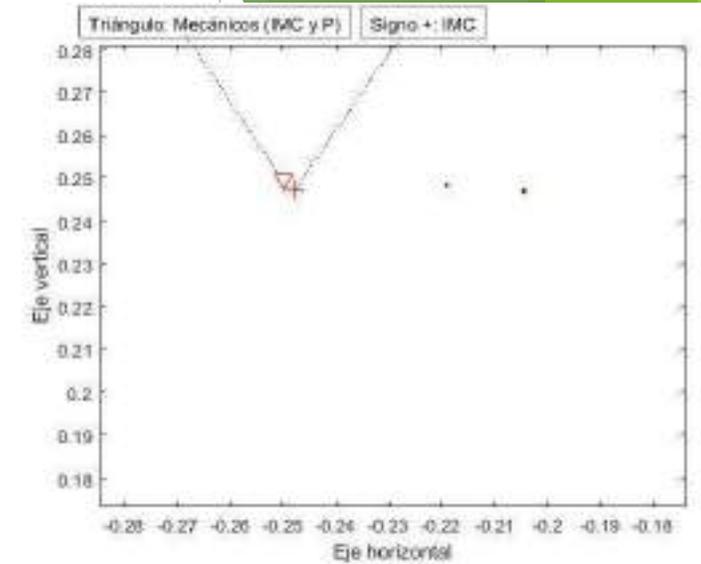
Estudio de los factores Mecánicos Grupo Minas



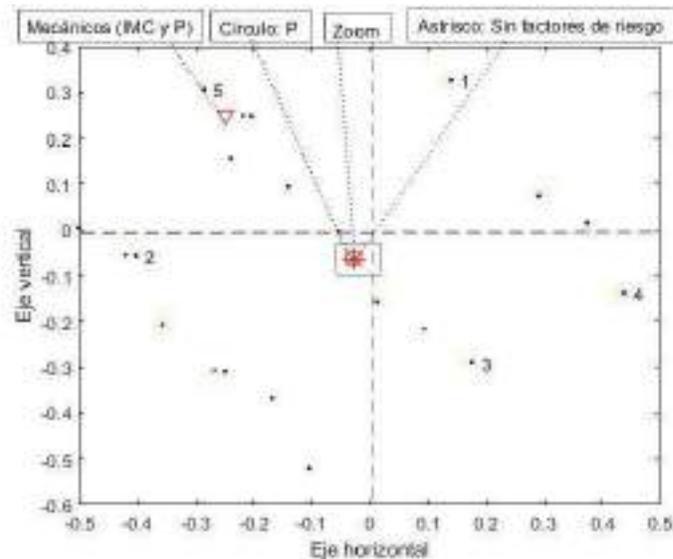
a) Factores Mecánicos



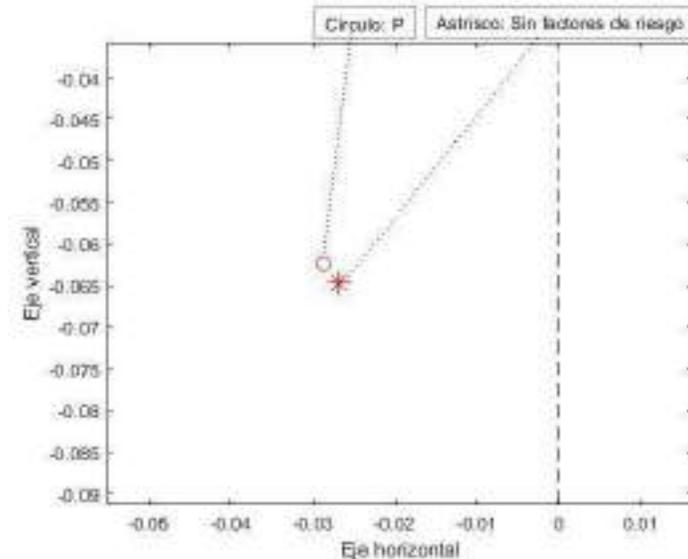
b) Se elimina Pre. Arterial



c) Ampliación de b

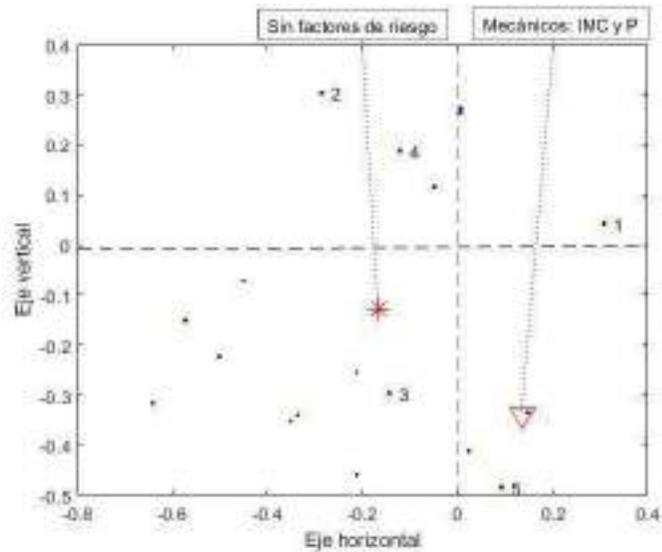


d) Se elimina IMC

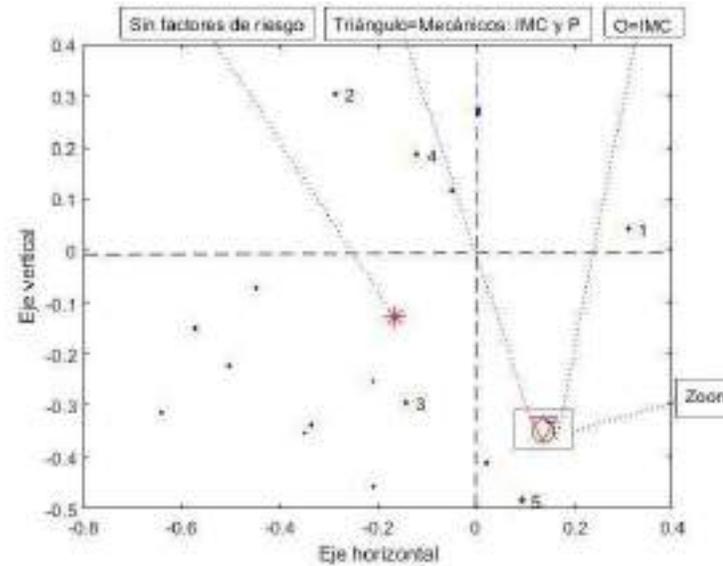


e) Apliación de d

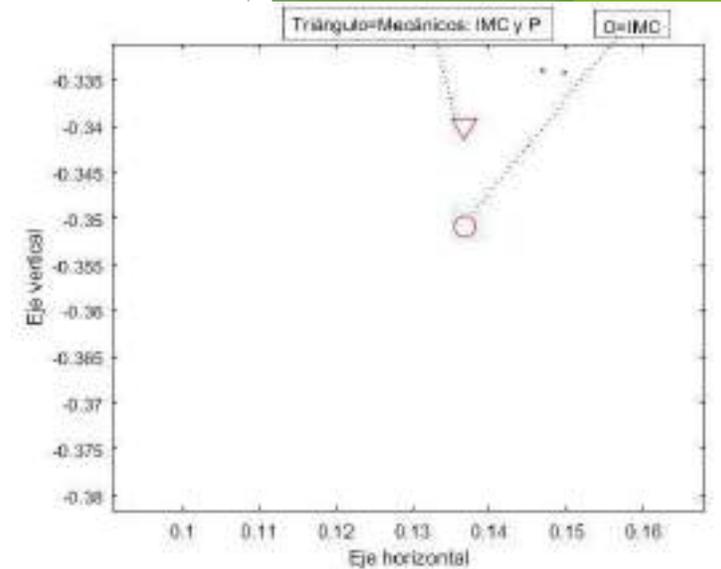
Estudio de los factores Mecánicos Grupo Finanzas



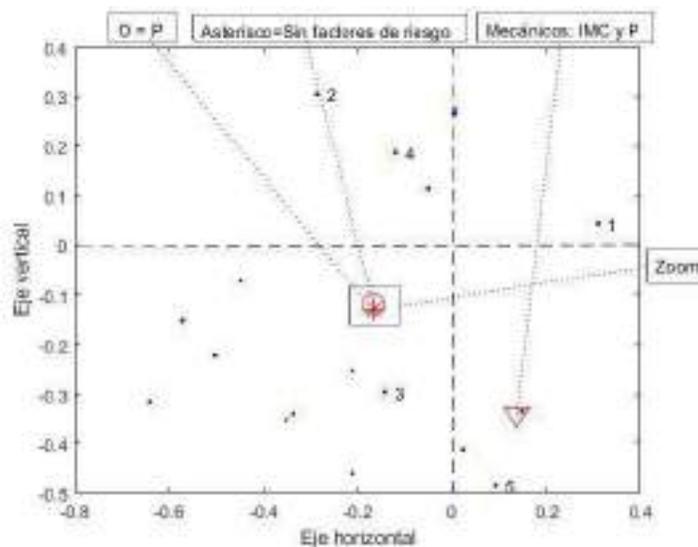
a) Factores Mecánicos



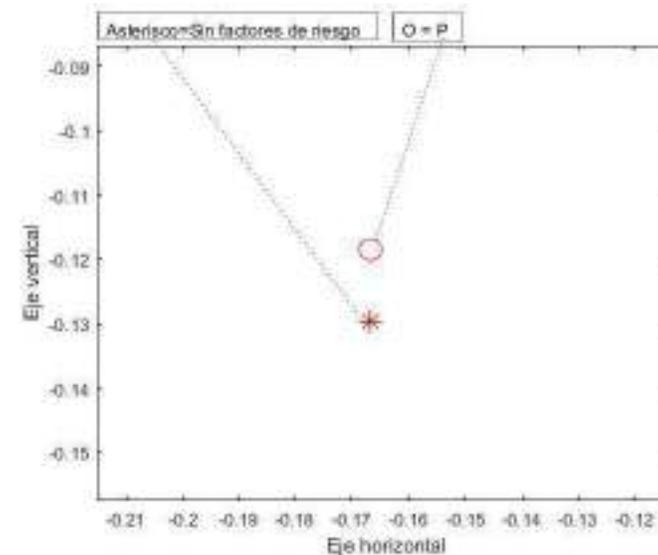
b) Se elimina Pre. Arterial



c) Ampliación de b

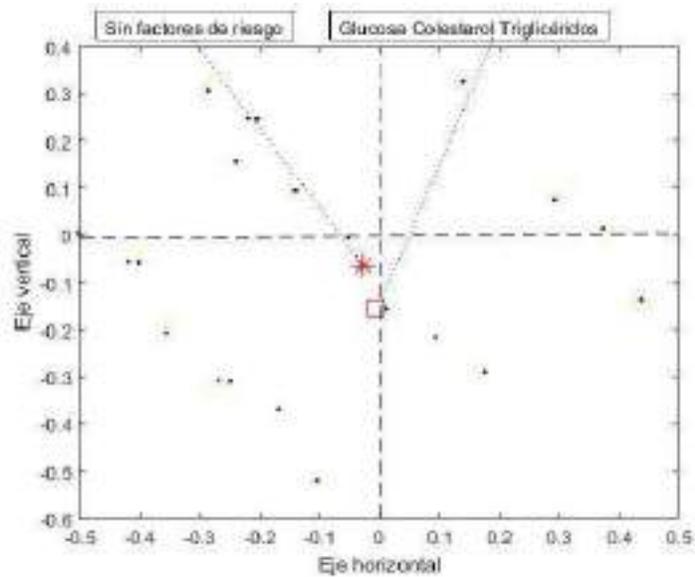


d) Se elimina IMC

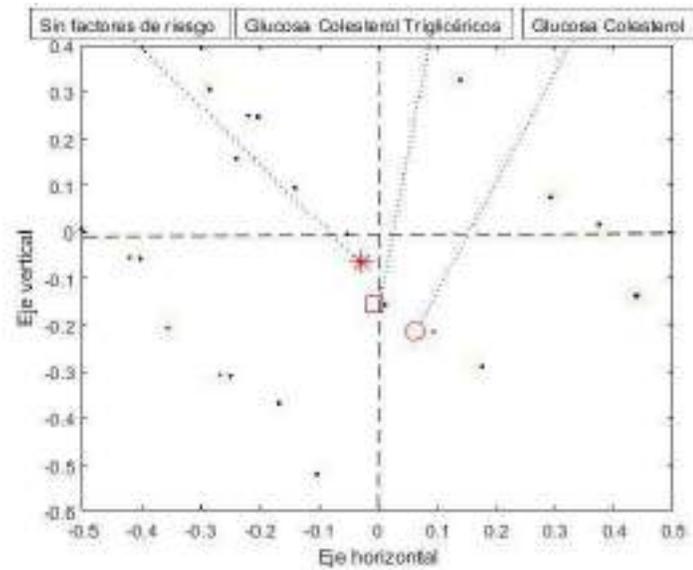


e) Apliación de d

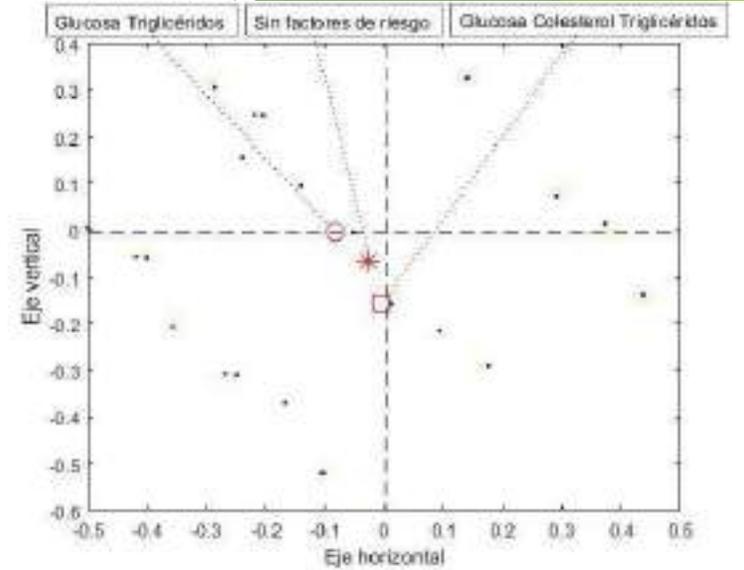
Estudio de los factores Químicos Grupo Minas



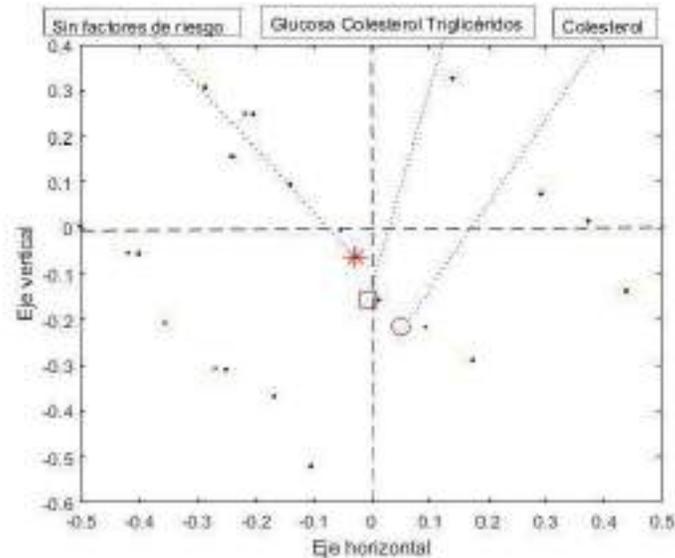
a) Factores Químicos



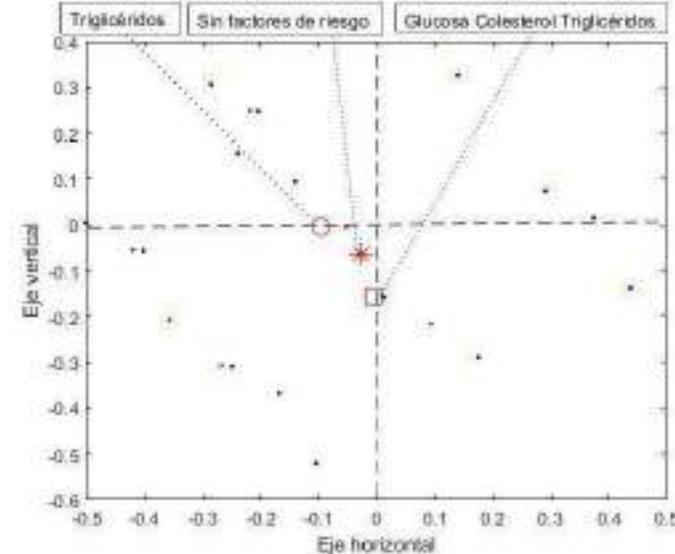
b) Se elimina Triglicéridos



c) Se elimina Colesterol

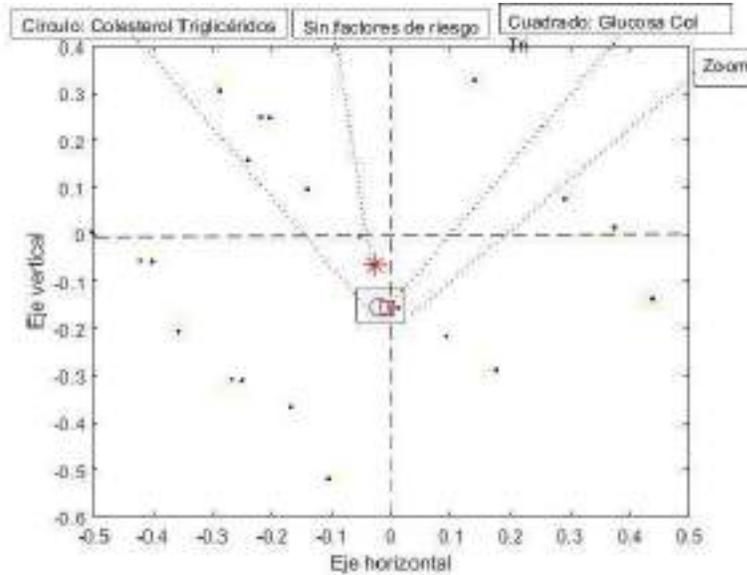


d) Se elimina Triglicéridos y

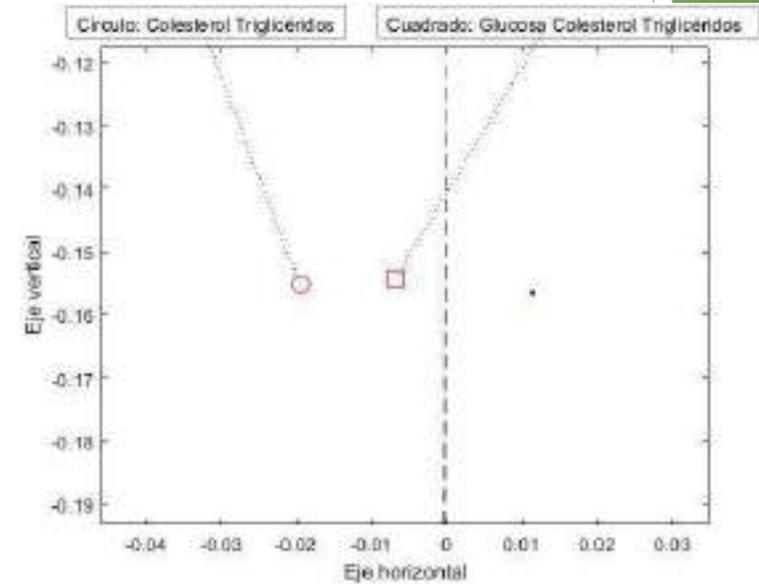


e) Se elimina Colesterol y Glucosa

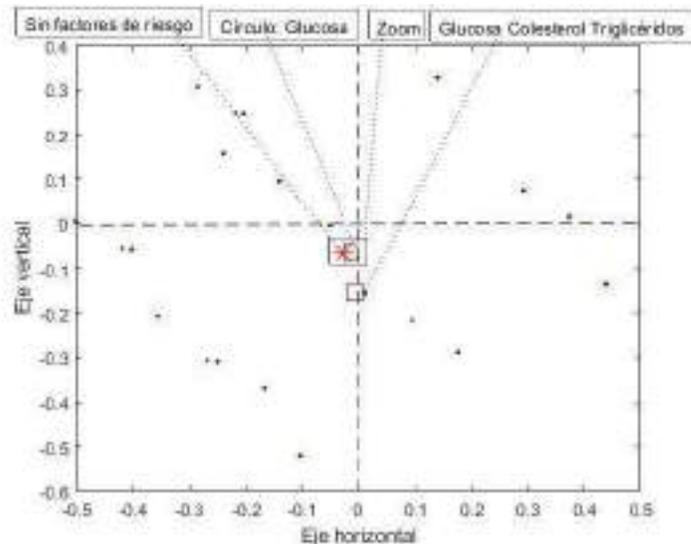
Estudio de los factores Químicos Grupo Minas (continuación)



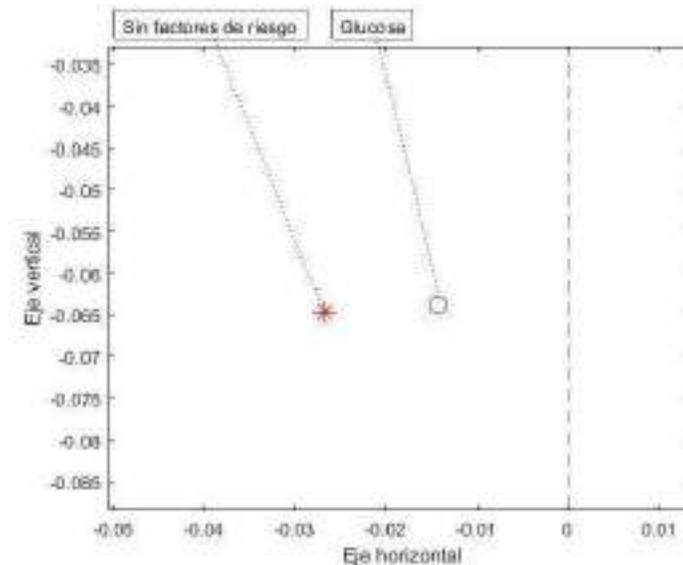
f) Se elimina Glucosa



g) Ampliación de f

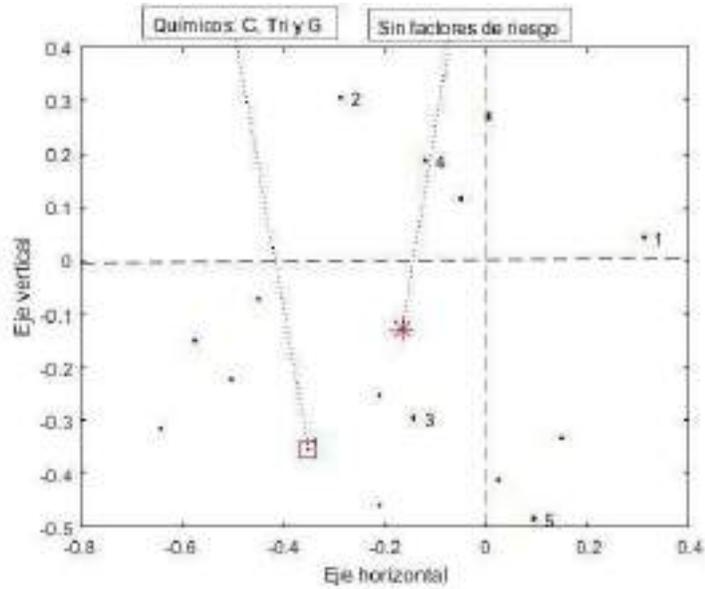


h) Se elimina Colesterol y Triglicéidos

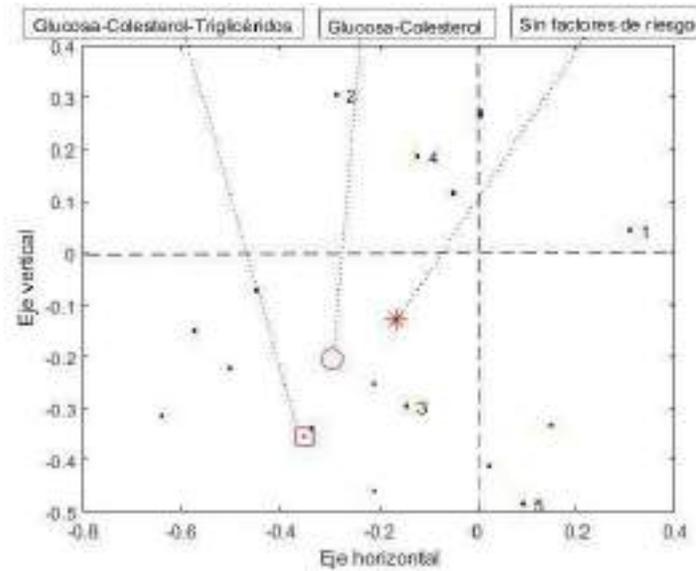


i) Ampliación de h

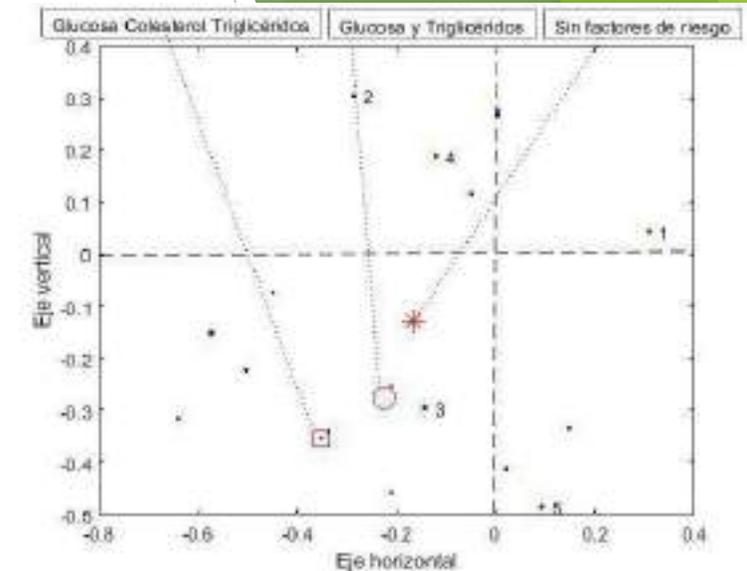
Estudio de los factores Químicos Grupo Finanzas



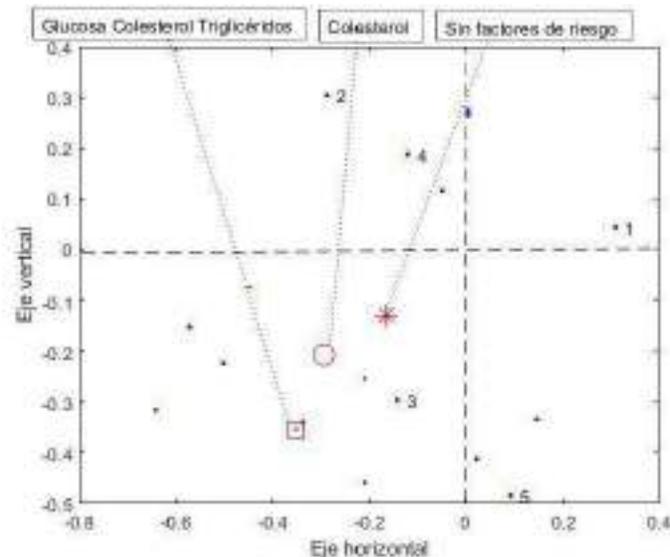
a) Factores Químicos



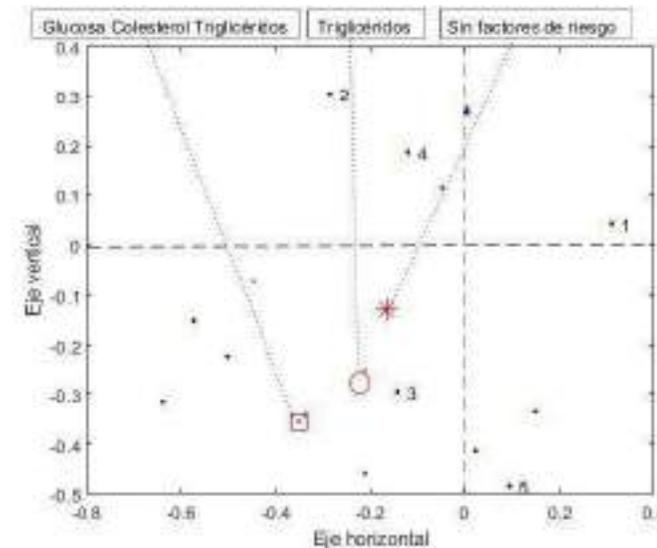
b) Se elimina Triglicéridos



c) Se elimina Colesterol

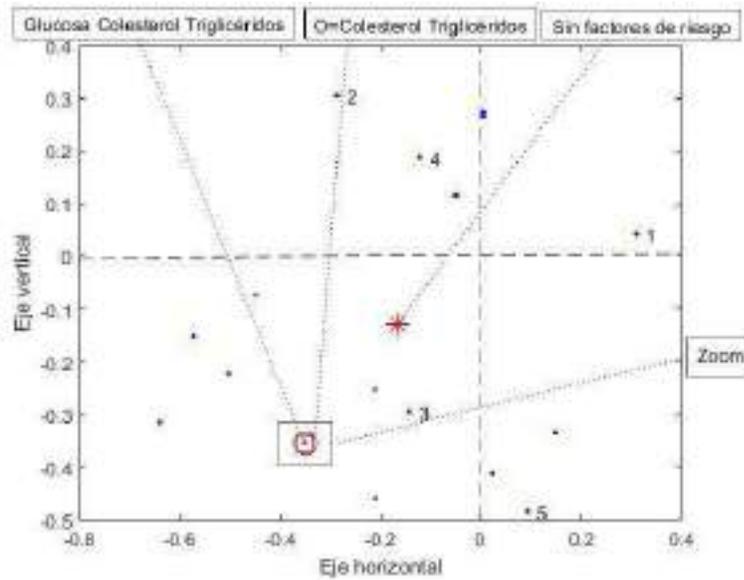


d) Se elimina Triglicéridos y

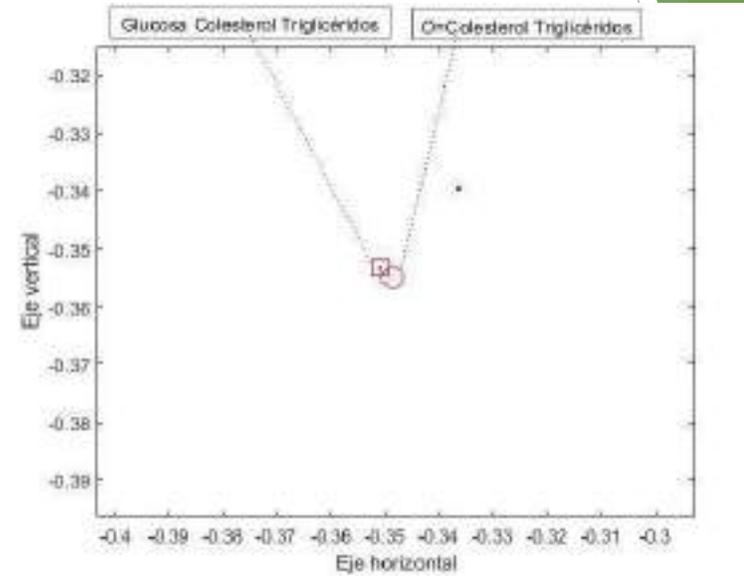


e) Se elimina Colesterol y Glucosa

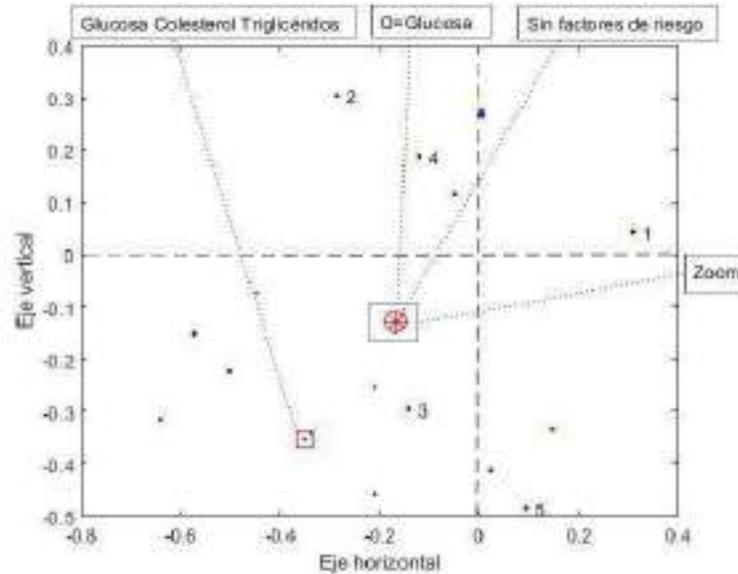
Estudio de los factores Químicos Grupo Finanzas (continuación)



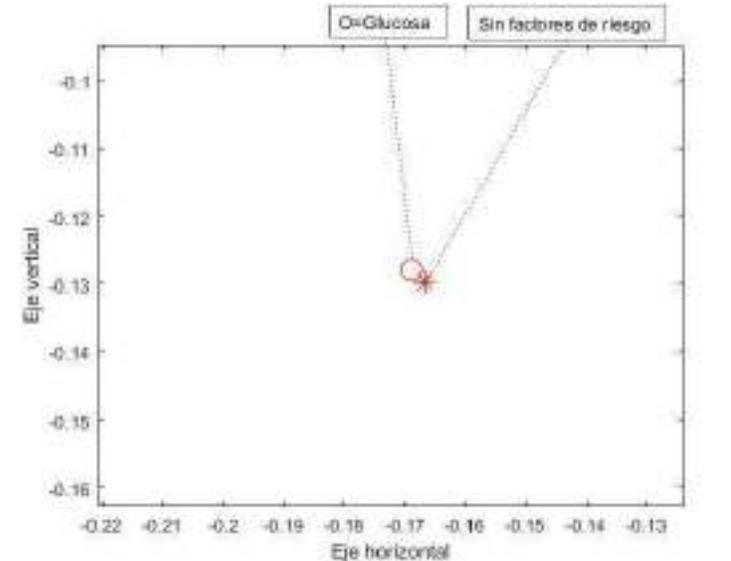
f) Se elimina Glucosa



g) Ampliación de f

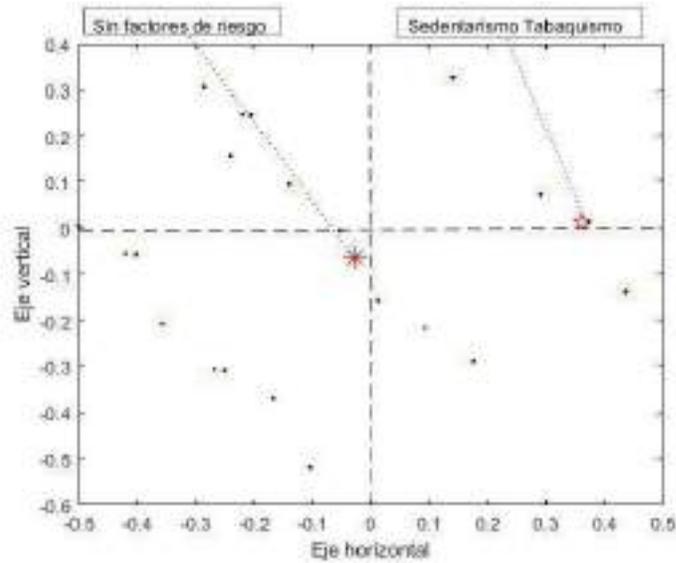


h) Se elimina Colesterol y Triglicéridos

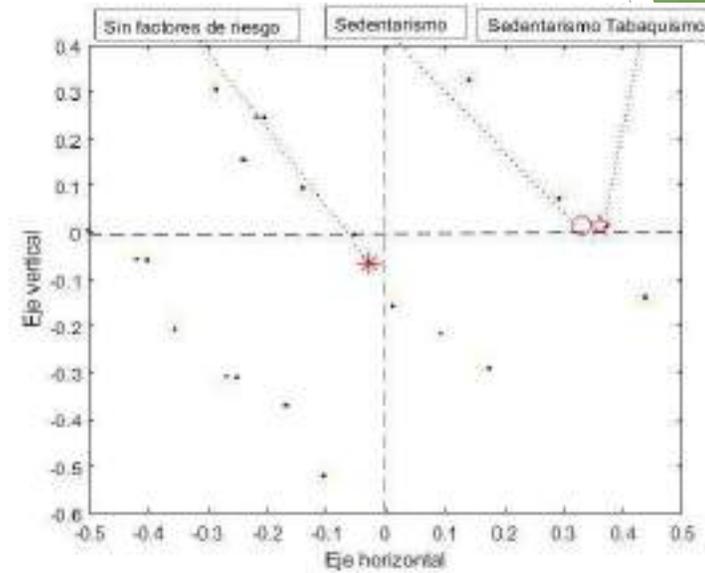


i) Ampliación de h

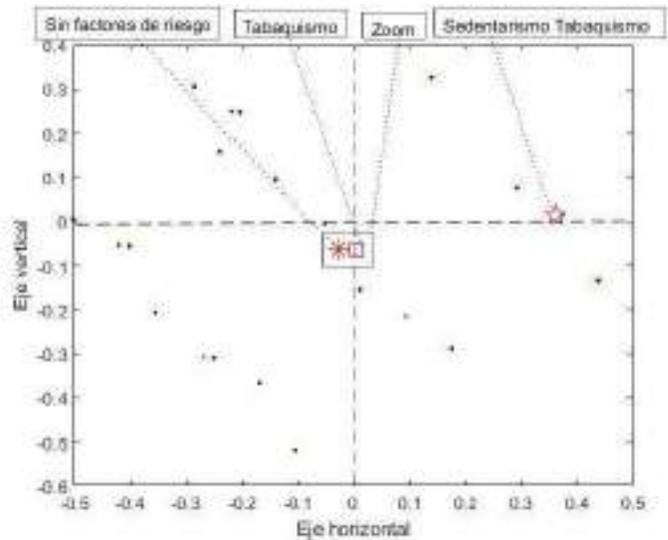
Estudio de los factores Hábitos Grupo Minas



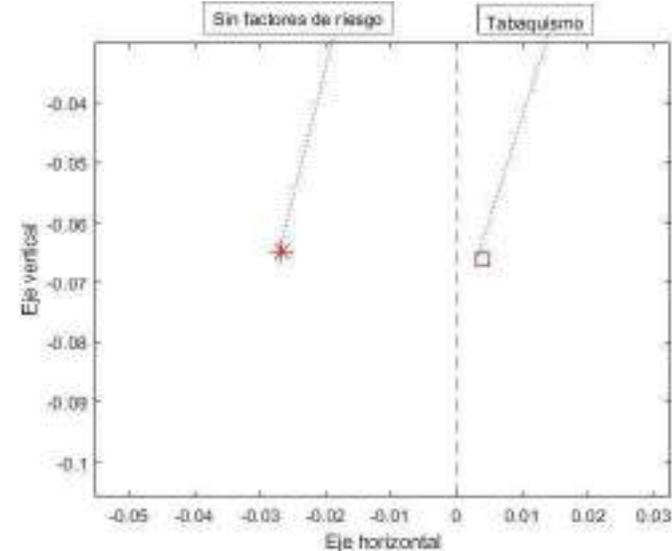
a) Factores Hábitos



b) Se elimina Tabaquismo

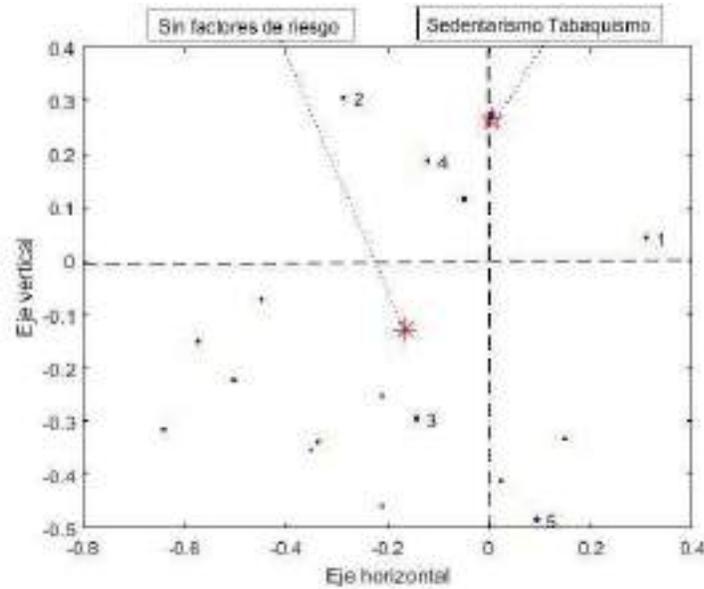


c) Se elimina Sedentarismo

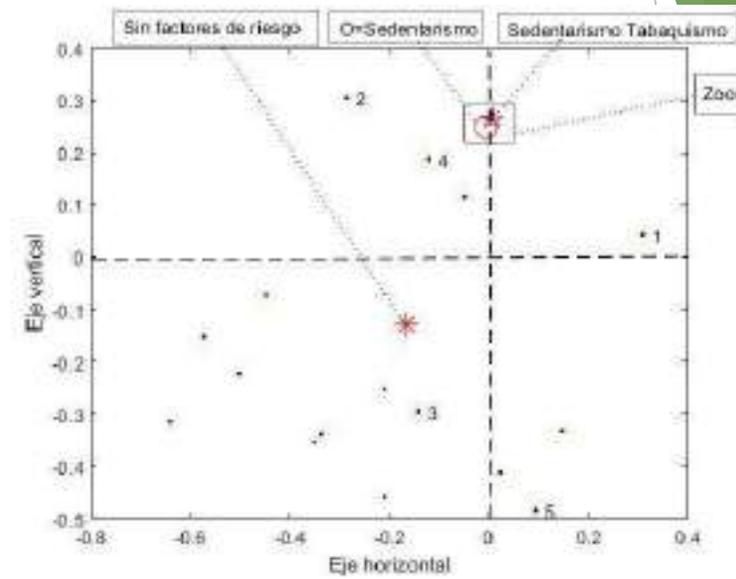


d) Ampliación de c

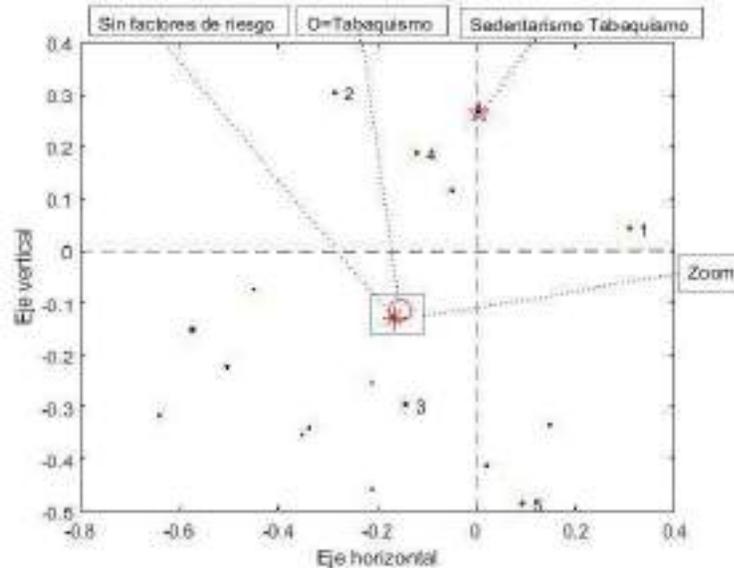
Estudio de los factores Hábitos Grupo Finanzas



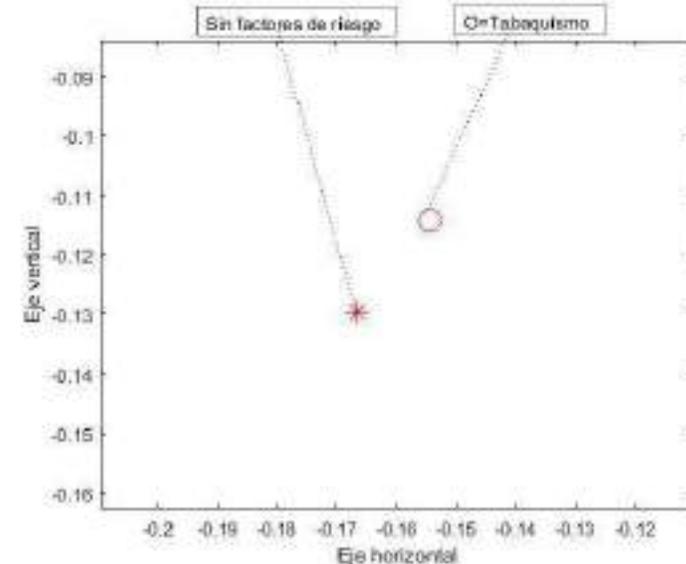
a) Factores Hábitos



b) Se elimina Tabaquismo

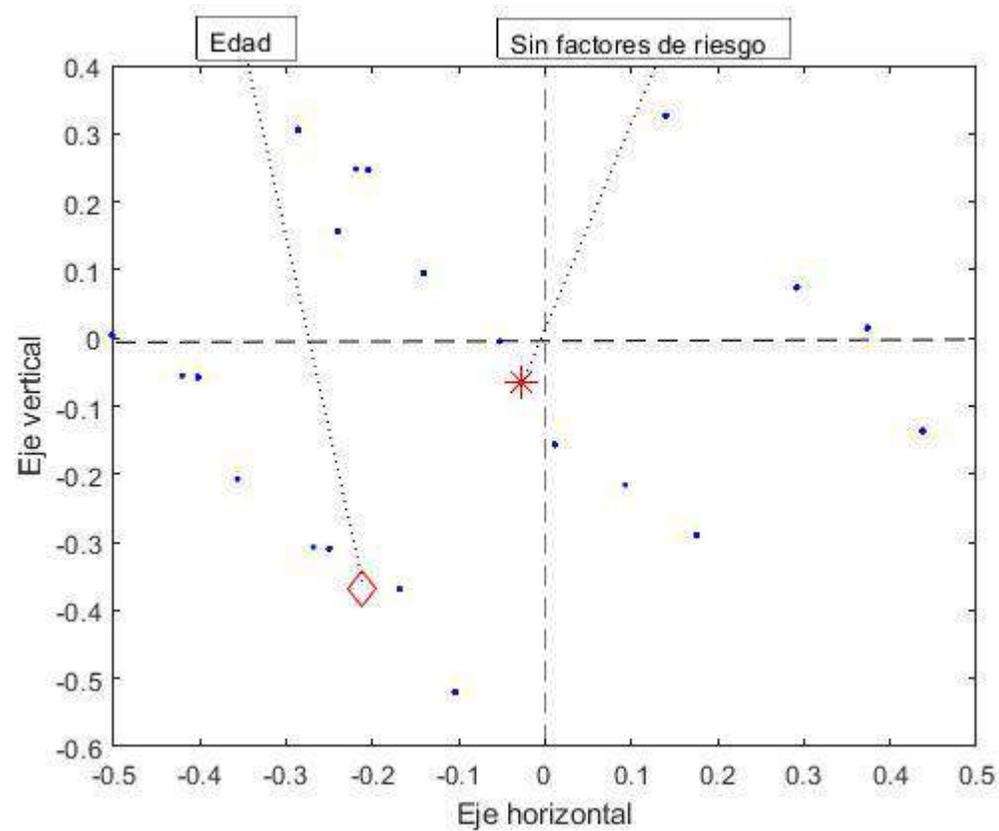


c) Se elimina Sedentarismo

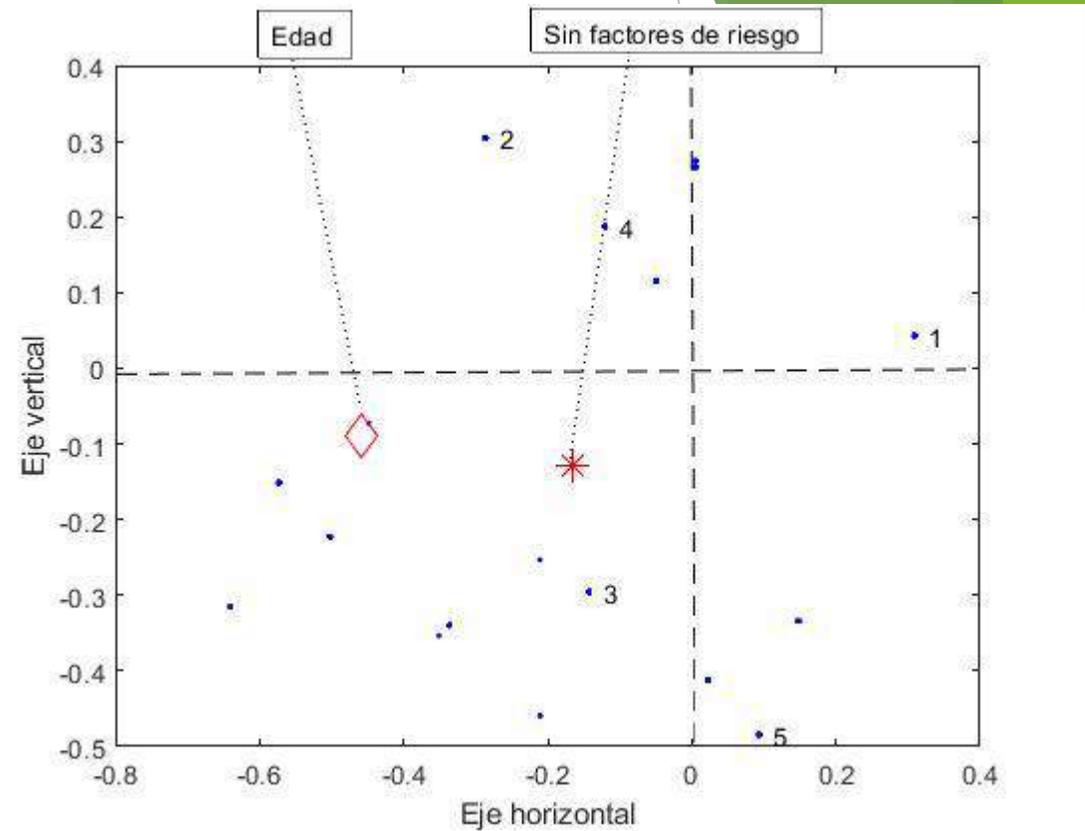


d) Ampliación de c

Ubicación de la variable Edad en los Grupos Minas y Finanzas



a) Variable Edad Grupo Minas

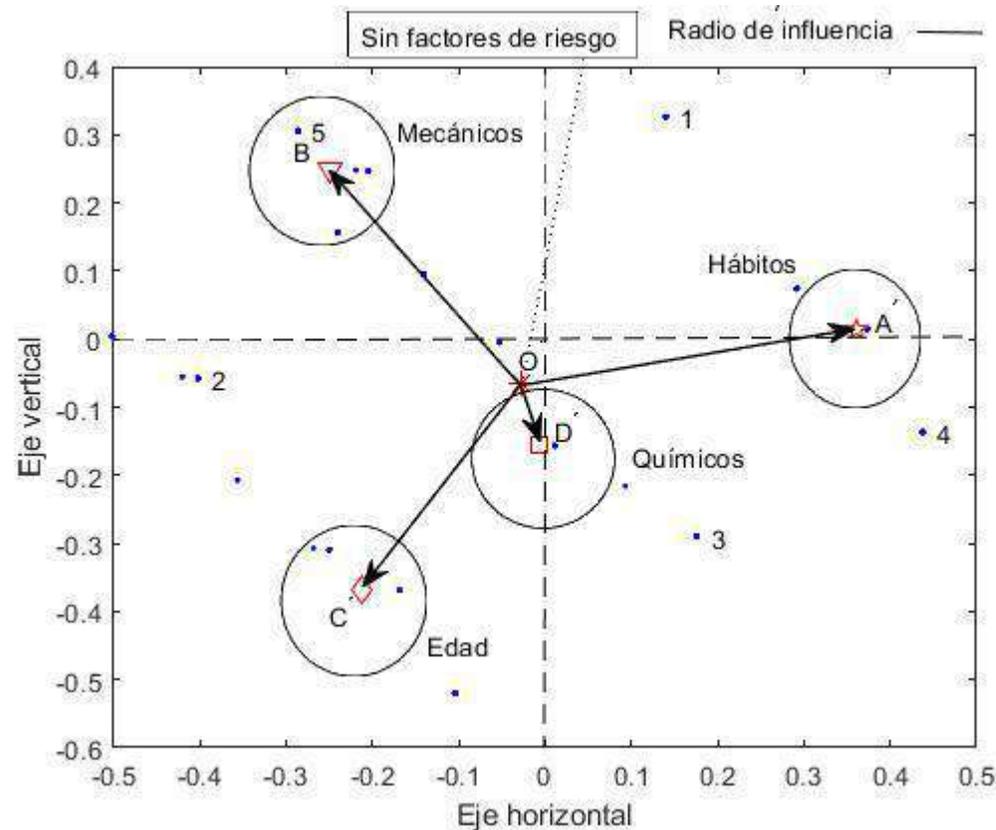


b) Variable Edad Grupo Finanzas

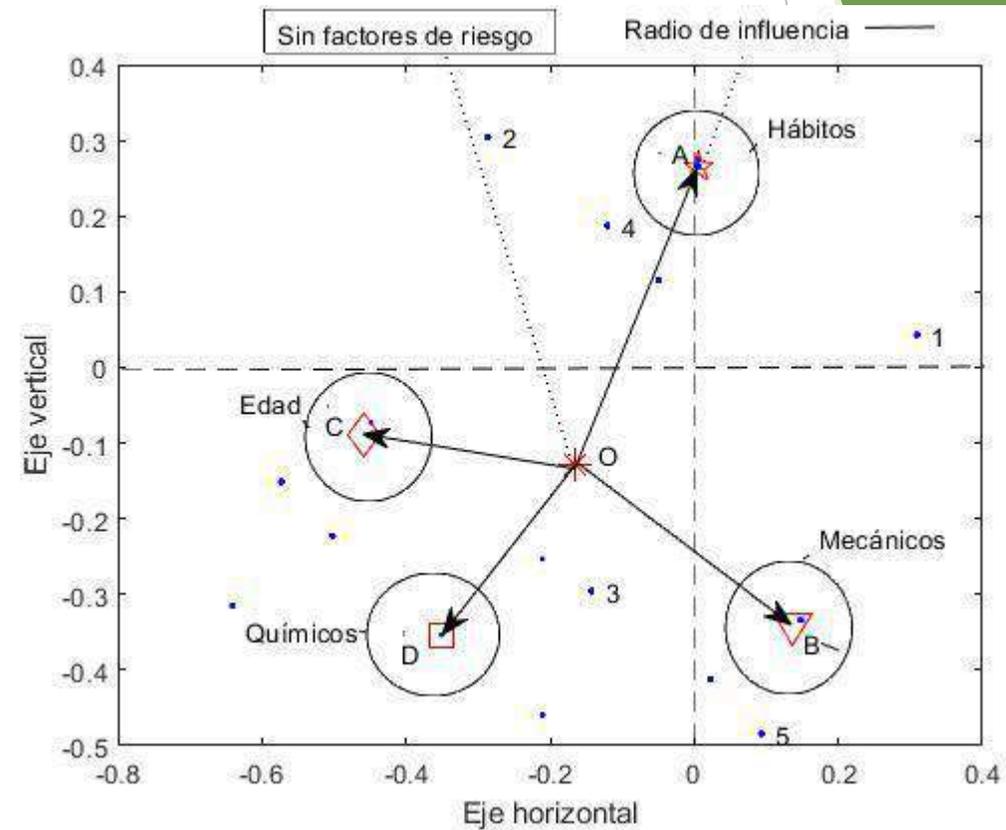
Análisis predictivo mediante los Factores de Riesgo

- ▶ Una situación deseable es poder predecir el estado de salud de un paciente en base a los resultados de sus exámenes médicos, resultados que con ayuda del escalado multidimensional se pueden representar como un punto en el plano de similitudes, generado por un grupo determinado; por tanto, la ubicación de dicho punto nos dará información de posibles factores de riesgo presentes y sugerirá opciones para contrarrestarlos.
- ▶ Para poder predecir un estado de salud en función de la ubicación del paciente en el gráfico de similitudes, es necesario conocer la zona de influencia de cada uno de los factores de riesgo, zona de influencia que queda definida por la ubicación del extremo del vector de cada factor de riesgo.
- ▶ La representación de un paciente con cierto factor de riesgo, por medio de un vector con el origen en la observación sin factores de riesgo, por ejemplo el vector OA; define la zona de influencia de cada factor de riesgo, situación que la podemos visualizar en los círculos de las Figuras

Zonas de influencia de los factores: Mecánicos, Químicos, Hábitos y Edad

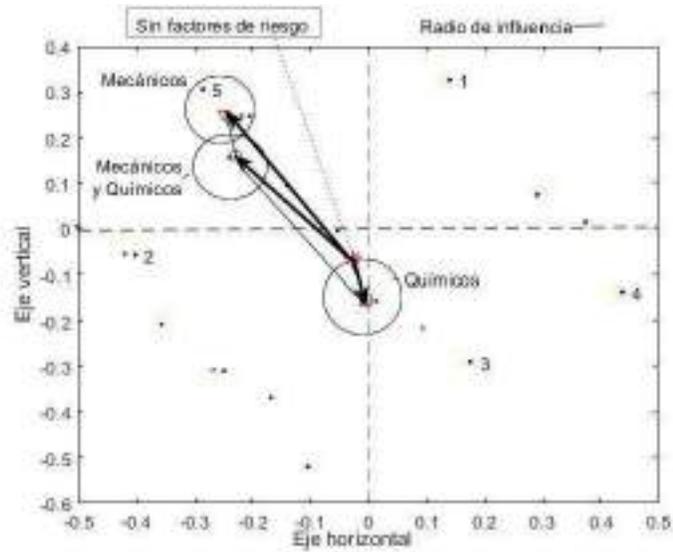


a) Zonas de influencia de los Factores para el Grupo Minas

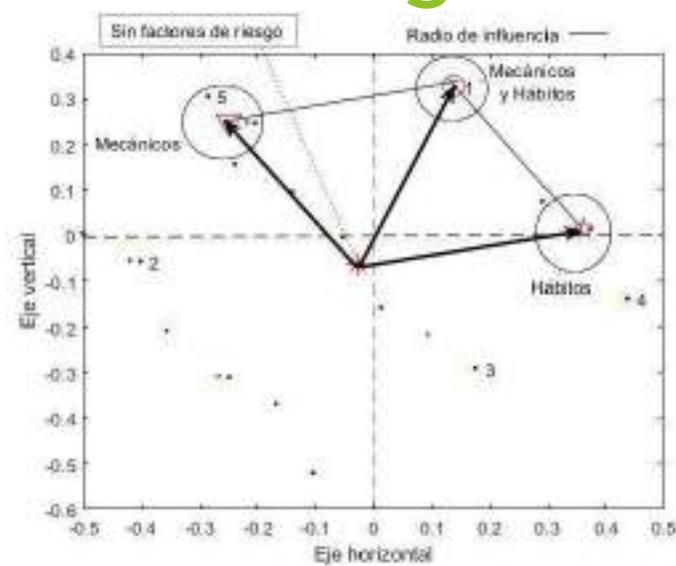


b) Zonas de influencia de los Factores para el Grupo Finanzas

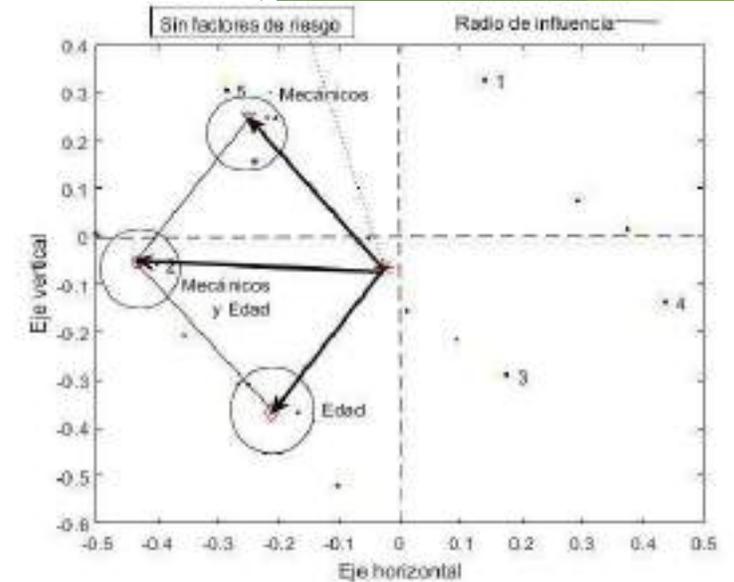
Suma de Factores de riesgo Grupo Minas



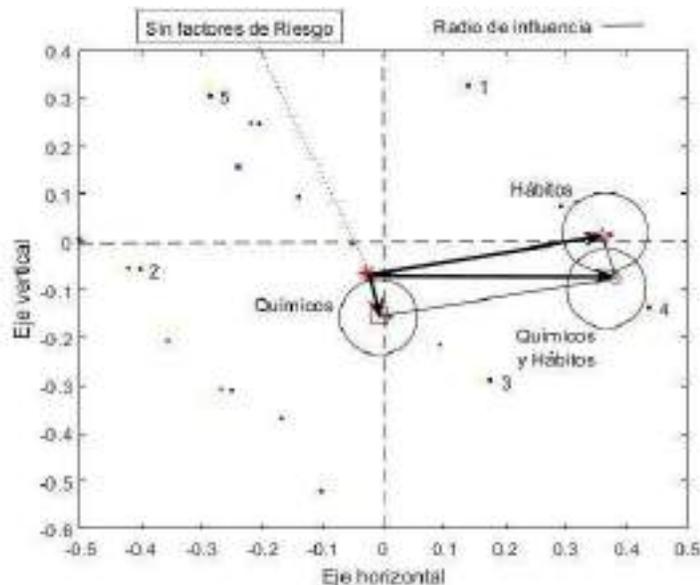
a) Mecánicos y Químicos



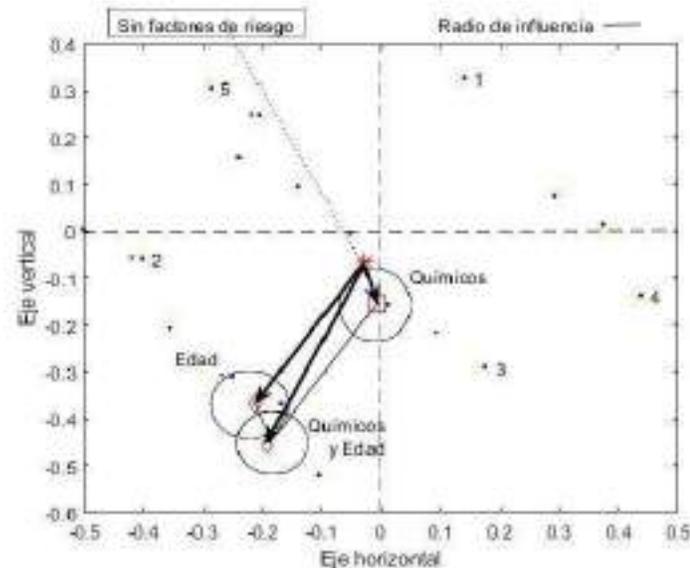
b) Mecánicos y Hábitos



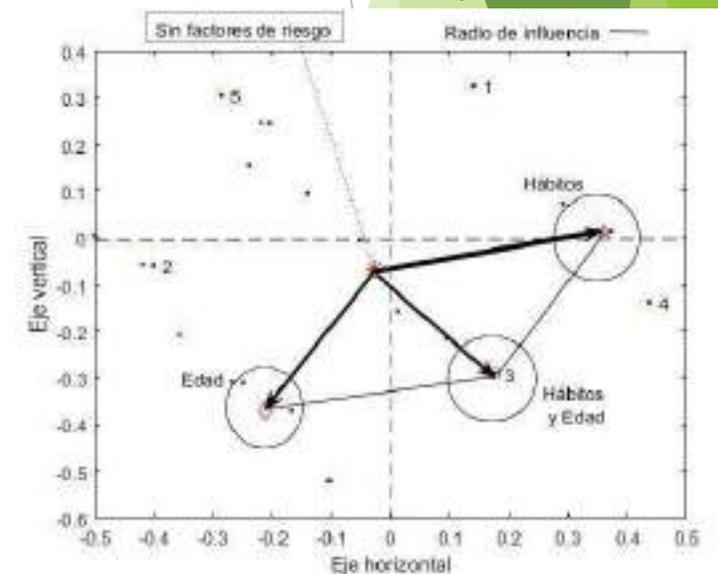
c) Mecánicos y Edad



d) Hábitos y Químicos

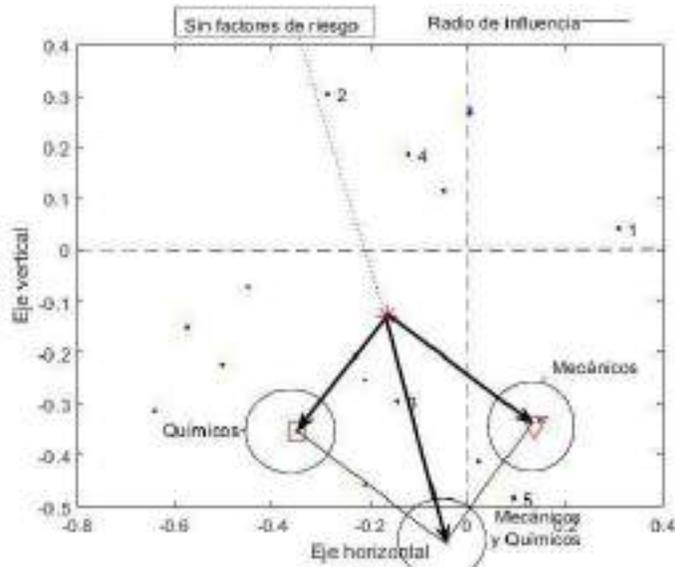


e) Edad y Químicos

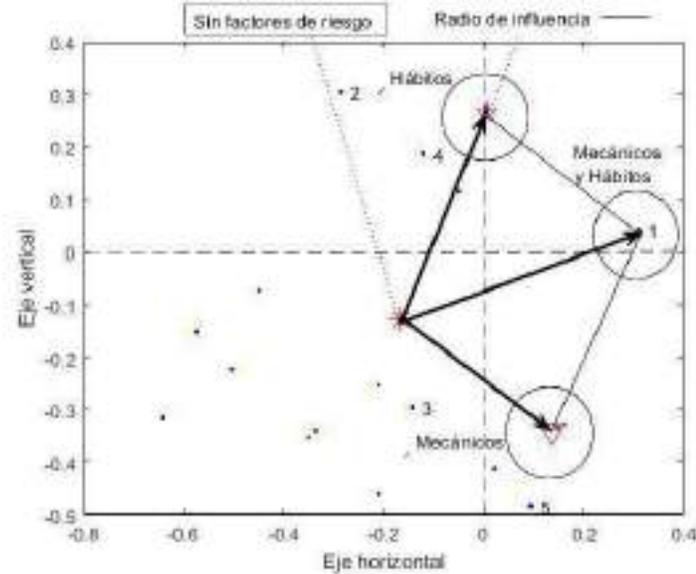


f) Edad y Hábitos

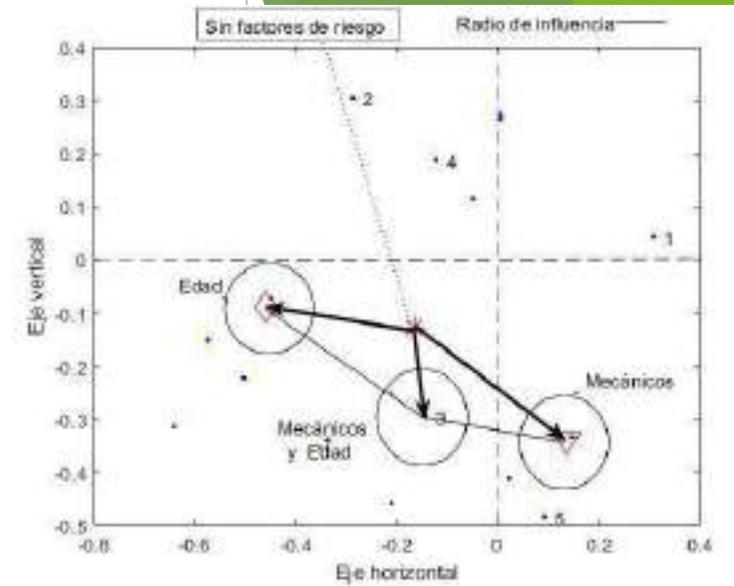
Suma de Factores de riesgo Grupo Finanzas



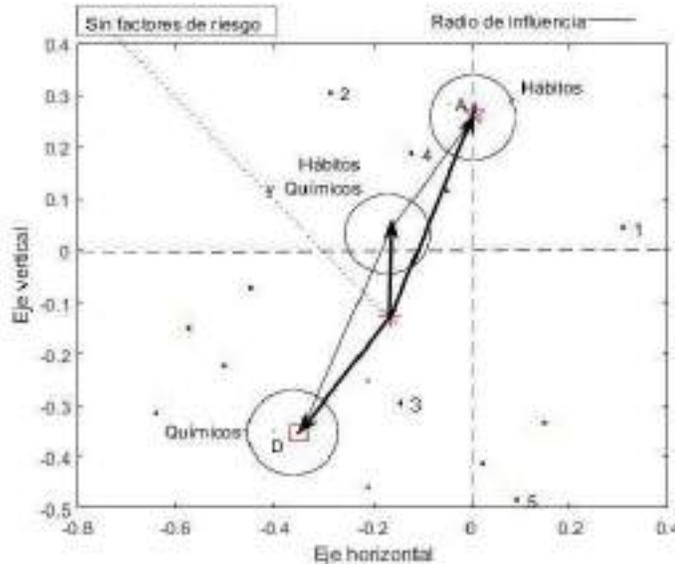
a) Mecánicos y Químicos



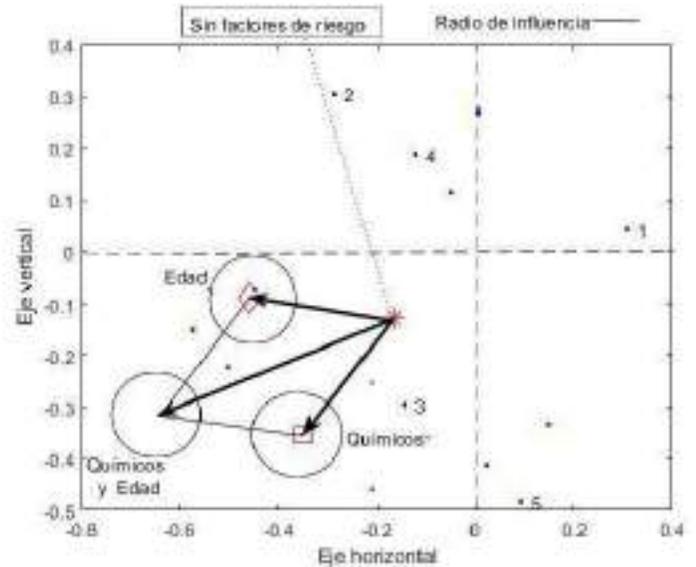
b) Mecánicos y Hábitos



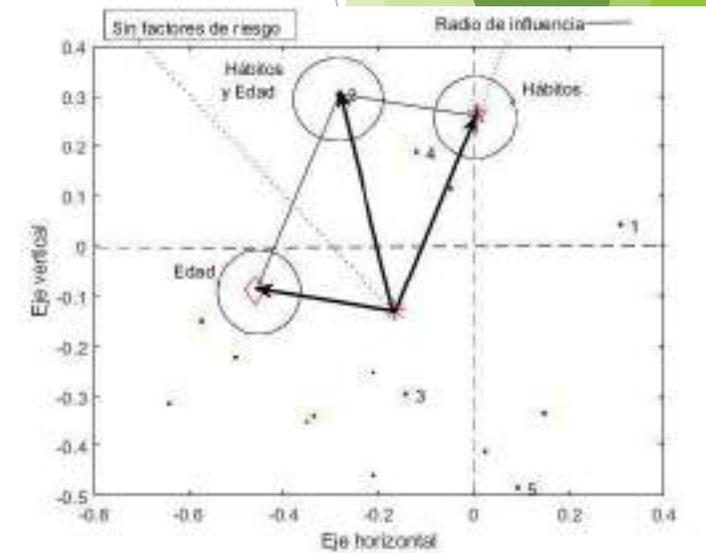
c) Mecánicos y Edad



d) Hábitos y Químicos



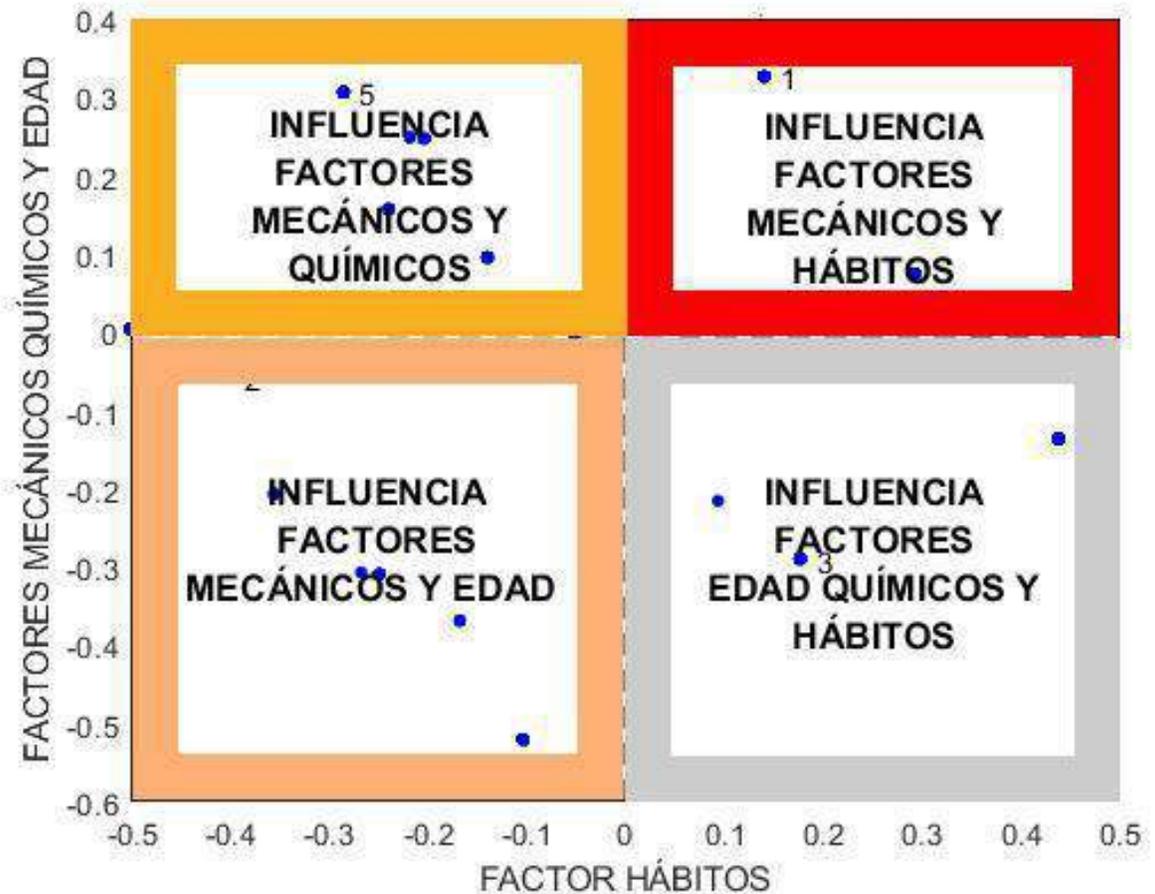
e) Edad y Químicos



f) Edad y Hábitos

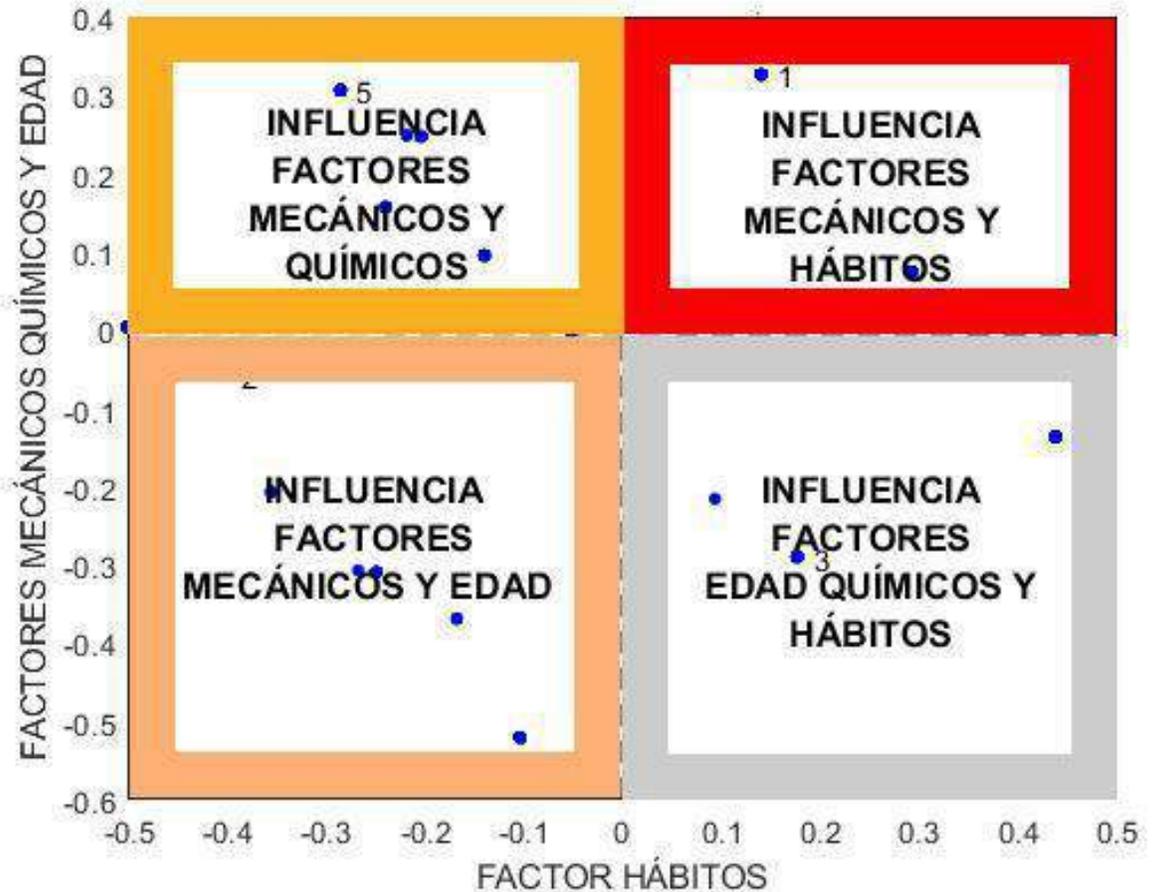
Influencia de los diversos factores de riesgo para Explotación de Minas

Al sobreponer los gráficos de las Figuras de las zonas de influencia de uno y de dos factores de riesgo se obtiene un nuevo gráfico que indica los factores de riesgo para los diferentes cuadrantes del plano del grupo Explotación de Minas.



Influencia de los diversos factores de riesgo para Actividades Financieras

Al sobreponer los gráficos de las Figuras de las zonas de influencia de uno y de dos factores de riesgo se obtiene un nuevo gráfico que indica los factores de riesgo para los diferentes cuadrantes del plano del grupo Actividades Financieras.



Conclusiones

Conclusiones del Análisis Descriptivo

Las características de la población objeto del presente estudio son:

- ▶ La población tiende al sobrepeso, aproximadamente 2 de cada 3 personas sobrepasan los límites de normalidad de la variable IMC.
- ▶ El 39.10 % de la población tiene Colesterol alto.
- ▶ El 36.46 % de la población tiene Triglicéridos alto.
- ▶ El 14.10 % de la población tiene Glucosa alta.
- ▶ El 3.00 % de la población tiene Presión Arterial alta.
- ▶ El 90.00 % de la población está en el rango [18, 50] años.

Los datos sobre IMC, Colesterol y Triglicéridos indican una alimentación no adecuada en un alto porcentaje de la población.

Conclusiones del Análisis por Índice de Salud

- ▶ El Índice de Salud proporciona información del estado de salud de los pacientes de la data respecto al Síndrome Metabólico.
- ▶ La distribución de la población de estudio en las categorías definidas por este índice es:

Bajo nivel de normalidad, 695 pacientes (5.62 %).

Normal, 4 536 pacientes (36.69 %).

Sobre nivel de normalidad, 6 244 pacientes (50.51 %).

Crítico, 888 pacientes (7.18 %).

Los resultados presentan cierto grado de alarma, ya que el 58 % de la población se encuentra fuera del rango de normalidad.

Conclusiones del Análisis por Índice de Salud

la distribución de la población de acuerdo al Estado de Salud, para los grupos Explotación de Minas y Actividades Financieras es:

Para Explotación de Minas: Bajo 142 pacientes (4.27 %), Normal 996 pacientes (29.92 %), Sobre 1 869 pacientes (56.14 %) y Crítico 323 pacientes (9.67 %).

Para Actividades Financieras: Bajo 112 pacientes (5.09 %), Normal 968 pacientes (44.03 %), Sobre 1 021 pacientes (46.43%) y Crítico 98 pacientes (4.45 %).

Comparando la distribución de las dos poblaciones se observa que la actividad económica Explotación de Minas posee mayor porcentaje de pacientes en los Estado de Salud Sobre y Crítico, en comparación con Actividades Financieras.

Conclusiones del Análisis por factores de riesgo

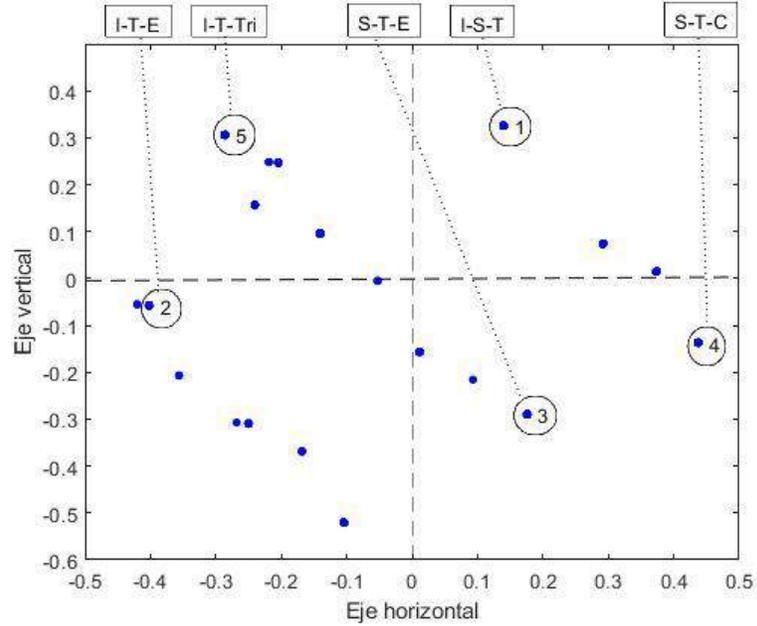
- ▶ Este análisis permite categorizar a la población según el número de factores de riesgo, de acuerdo a la literatura médica únicamente la población de los grupos C0, C1 y C2 tienen una condición de salud buena, que corresponden a 3191 pacientes (25.81 %); el resto presenta diversas condiciones de riesgo para su salud.
- ▶ En el Grupo Explotación de Minas el porcentaje de pacientes en condición de buena salud es 17.93 %, mientras que en el grupo Actividades Financieras el porcentaje es 29.45 %; lo que indica una mejor condición de salud en este último grupo.

Conclusiones del Análisis Multidimensional

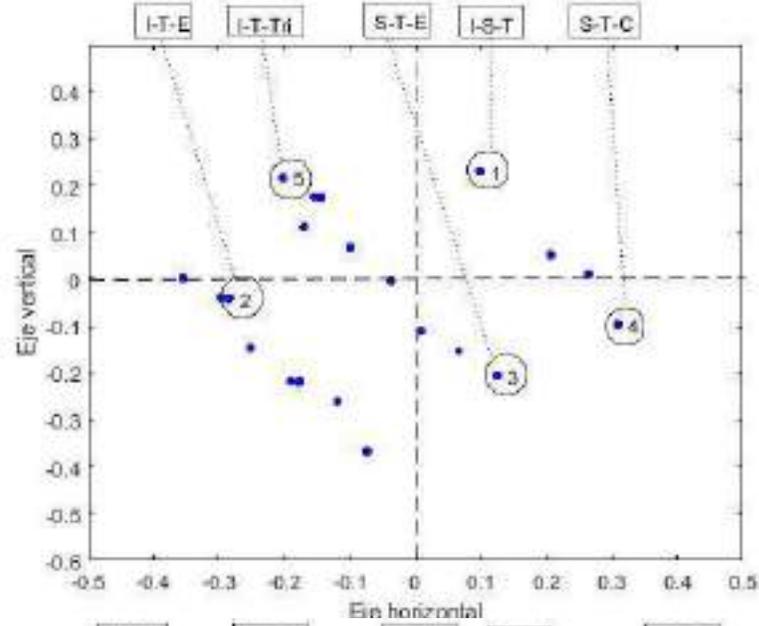
- ▶ El análisis es robusto respecto al uso de diversos coeficientes de similitud , ya que al aplicar al grupo Explotación de Minas con tres factores críticos, los coeficientes de similitud de Sokal-Michener, Rogers-Tanimoto y Russel-Rao se obtienen resultados similares.

Similaridades con:

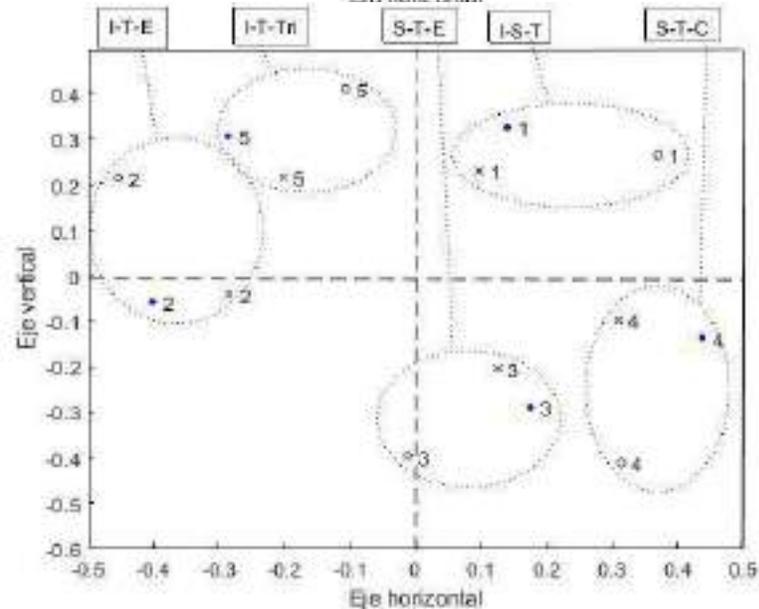
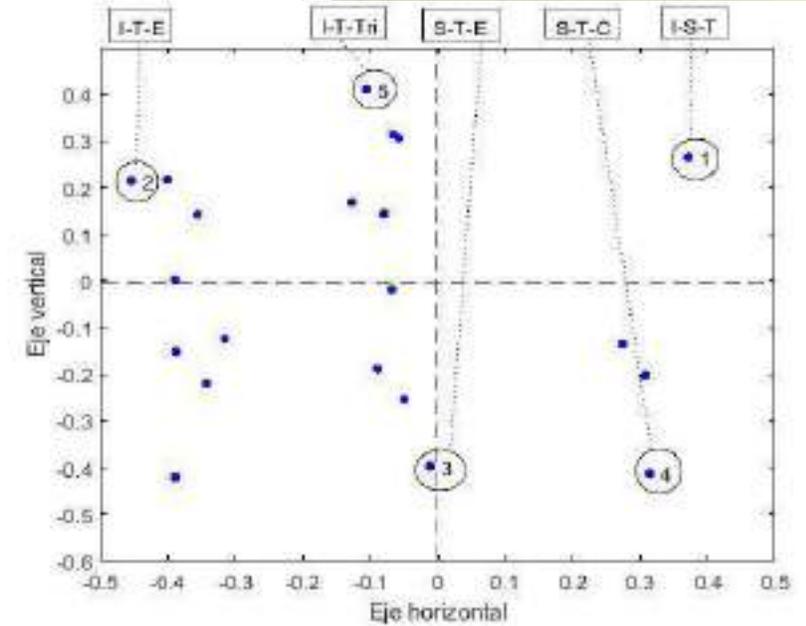
Sokal-Michener



Russell-Rao



Rogers-Tanimoto



- Sokal-Michener
- o Russell-Rao
- x Rogers-Tanimoto

Factores de riesgo grupo Explotación de Minas usando:

Sokal-Michener

Observación	Factor 1	Factor 2	Factor 3	Porcentaje
1	IMC	Sedentarismo	Tabaquismo	24
2	IMC	Tabaquismo	Edad	17
3	Sedentarismo	Tabaquismo	Edad	15
4	Colesterol	Sedentarismo	Tabaquismo	10
5	IMC	Triglicéridos	Tabaquismo	5

Russell-Rao

Observación	Factor 1	Factor 2	Factor 3	Porcentaje
1	IMC	Sedentarismo	Tabaquismo	24
2	IMC	Tabaquismo	Edad	17
3	Sedentarismo	Tabaquismo	Edad	15
4	Colesterol	Sedentarismo	Tabaquismo	10
5	IMC	Triglicéridos	Tabaquismo	8

Rogers-Tanimoto

Observación	Factor 1	Factor 2	Factor 3	Porcentaje
1	IMC	Sedentarismo	Tabaquismo	24
2	IMC	Tabaquismo	Edad	17
3	Sedentarismo	Tabaquismo	Edad	15
4	Colesterol	Sedentarismo	Tabaquismo	10
5	IMC	Triglicéridos	Tabaquismo	8

Conclusiones del Análisis Multidimensional (continuación)

- ▶ En lo referente a los factores Mecánicos, que incluyen las variables IMC y Presión Arterial, resulta más eficiente realizar un control de la variable IMC para alcanzar un estado óptimo de salud.
- ▶ Los factores Químicos son: Glucosa, Colesterol y Triglicéridos, para un paciente que tenga estos factores de riesgo, la simulación indica que para lograr un estado de salud óptimo resulta eficiente controlar las variables Triglicéridos y Colesterol.
- ▶ Si el paciente tiene factores de riesgo Hábitos: Sedentarismo y Tabaquismo; lo conveniente para lograr un estado de salud óptimo, es controlar la variable Sedentarismo.
- ▶ La ubicación de los factores Mecánicos, Químicos, Hábitos y Edad; considerados en los gráficos de similaridades, de los grupos Explotación de Minas y Actividades Financieras, se diferencian debido a que los factores de riesgo presentan diferentes porcentajes en cada una de estas actividades.

Comentarios del análisis multidimensional (continuación)

- ▶ El vector que representa los factores Mecánicos en el grupo Explotación de Minas es de mayor longitud a su similar del grupo Actividades Financieras, lo que se puede interpretar que los factores Mecánicos son más importantes en el grupo Explotación de Minas.
- ▶ El vector que representa los factores Químicos en el grupo de Actividades Financieras es de mayor longitud a su similar del grupo Explotación de Minas, lo que indica que los factores Químicos son más relevantes en Actividades Financieras.
- ▶ Los vectores que representan a los factores Edad y Hábitos son de similar longitud, lo que indica que estos factores tienen igual influencia en los dos grupos.