**INCIDENCIA Y CARACTERÍSTICAS DE LAS LESIONES OCURRIDAS EN EL PERIODO DE FEBRERO A MAYO DEL AÑO 2017 EN LA CANTERA DE UN CLUB DE FÚTBOL DEL SUROCCIDENTE COLOMBIANO**

**Metodología**

Con el propósito de realizar un análisis exploratorio para determinar las características y naturaleza de las lesiones ocurridas en la población a analizar se propone el siguiente esquema de análisis para dar respuesta al objetivo propuesto:

1. Análisis descriptivo de la información
2. Análisis de correlación tanto para variables cualitativas como cuantitativas
3. Análisis de Correspondencias Múltiple

**Análisis de correspondencias múltiples (MCA, *Multiple Correspondence Analysis*)**

El Análisis de Correspondencias Múltiple es una técnica exploratoria multivariante que permite explorar y analizar la estructura de interrelaciones para individuos y variables, la cual tiene como insumo de entrada una matriz de variables categóricas en escala nominal. El objetivo principal del MCA consiste en estudiar variables e individuos en un espacio de menor dimensión con la menor perdida de información posible, de tal modo que sus relaciones (variables e individuos) se representen en términos de proximidad.

***Relaciones entre individuos***. Las relaciones entre individuos se miden en función de la cercanía, por tanto, individuos con el mismo patrón de respuesta, es decir, individuos que responden de un modo determinado las preguntas del instrumento/encuesta tienden a estar próximos.

***Relaciones entre variables***. En cuanto a las relaciones entre variables se deben distinguir los siguientes escenarios:

1. La proximidad entre niveles o categorías de respuesta de diferentes variables, indica que un amplio número de individuos tienden a seleccionar estos niveles conjuntamente por tanto se vuelve un patrón de respuesta predominante. Por el contrario, si existe un gran distanciamiento entre niveles o categorías de respuesta esto indica que es un patrón de respuesta poco común.
2. Aunque los niveles o categorías de respuesta de la misma variable no pueden seleccionarse de manera conjunta, su proximidad en el espacio de menor dimensión indica que estos niveles son en sí mismos similares ya que los individuos que las poseen presentan casi el mismo comportamiento en las otras variables.

La base fundamental del MCA sobre la cual se fundamenta consiste en la descomposición de la variabilidad o inercia total en la matriz de datos de variables cualitativas a partir de la construcción de nuevas variables o factores no correlacionados, denominados dimensiones, componentes o ejes principales. Estos nuevos componentes resumen la variación total en orden decreciente, es decir, la primera dimensión principal explica la mayor parte de la variabilidad de la tabla de datos, mientras la segunda dimensión explica parte de la variación no explicada por la primera. Finalmente la variación explicada por todas las dimensiones constituye el 100% de la inercia total de la tabla de datos.

Dado que las variables son categóricas, el cálculo de las proximidades entre variables e individuos parte del uso de la distancia Chi-cuadrado. Con esto los individuos y/o variables que muestran un patrón de respuesta promedio se encuentran en la coordenada del origen (punto 0,0 en el plano), mientras que los individuos que se alejan del origen muestran un patrón de respuesta con algunas particularidades a destacar, se ven caracterizados por las variables que muestran la mayor influencia en la composición de las dimensiones principales.

Para medir la influencia de las variables originales sobre las dimensiones principales estimadas, se hace uso de las contribuciones y el índice de calidad de representación, los cuales cuantifican la importancia de cada una de las modalidades o categorías de respuesta de las variables analizadas en la construcción de los dimensiones. Luego la interpretación de las dimensiones principales está en función de las modalidades de respuesta con las contribuciones/índice de calidad de representación más altas.

**Resultados**

En primera medida se realizó un análisis descriptivo de la información a partir del cálculo de frecuencias para las variables categóricas e indicadores de dispersión y tendencia central para las variables numéricas. Los resultados se presentan a continuación:

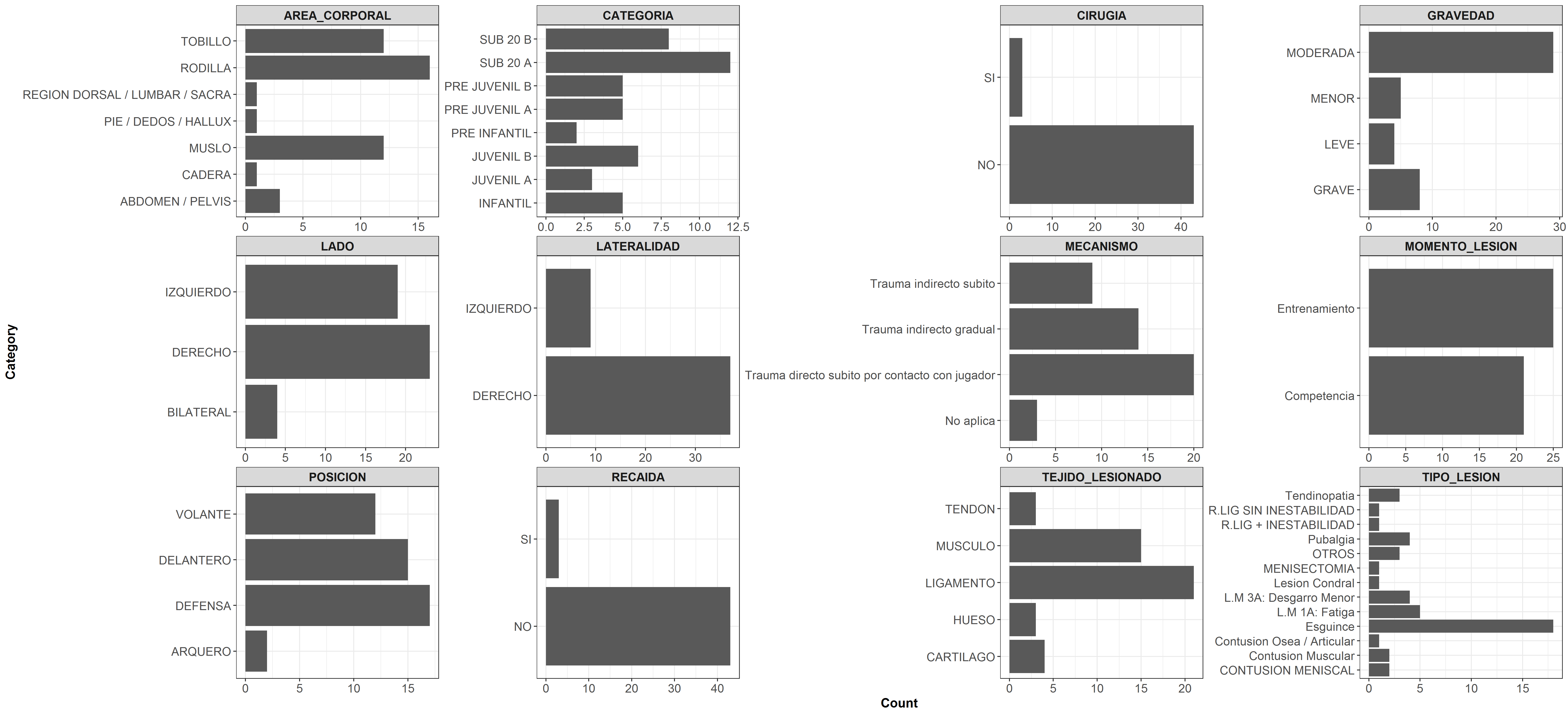


Figura . Gráficos de frecuencias - Variables cualitativas

Analizando la Figura 1, las áreas corporales con mayor número de lesiones en los jugadores evaluados son: tobillo, rodilla y muslo. En este sentido las posiciones con mayor número de lesiones corresponden a defensas, delanteros y volantes en orden decreciente. En cuanto a la gravedad de las lesiones prima el tipo moderado seguido del grave y el tipo de lesión más frecuente corresponde a los esguinces.

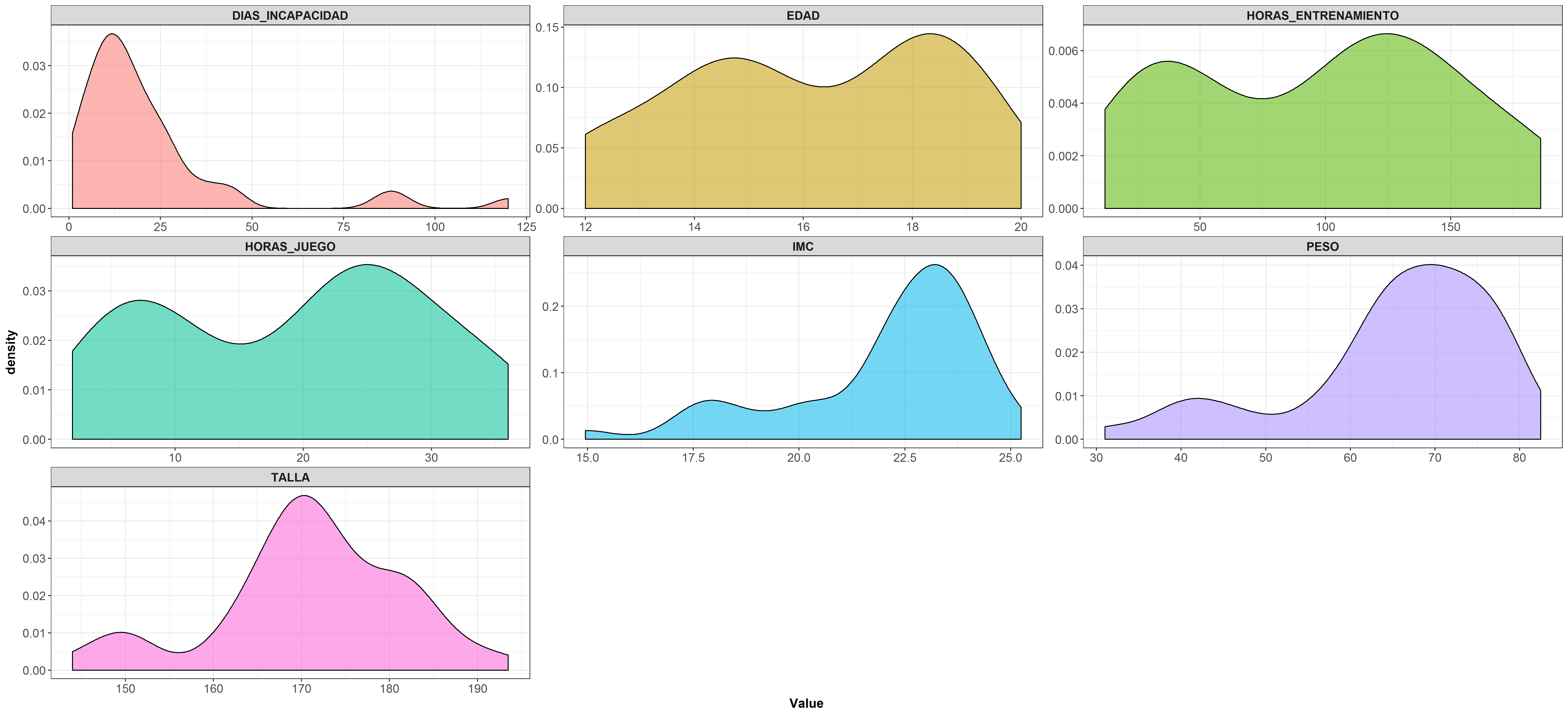


Figura . Gráficos de densidades - Variables cuantitativas

Por otro lado, realizando una exploración de las variables cuantitativas (ver Figura 2 y Tabla 1), se tiene un rango de edad que oscila de 12 a 20 años, con una edad promedio de 16 años. Estos jugadores presentan una talla promedio de 171 cm y un peso de medio de 65 kg. Adicionalmente presentan un IMC relativamente normal, no obstante destacándose los casos extremos de bajo peso (IMC = 14.95) y sobrepeso (IMC = 25.25).

Esta población de jugadores presenta una mediana de días de incapacidad de 16 días, cuyo rango principal va de 1 día a aproximadamente 30, sin embargo, hay registro de casos donde se sobrepasan los 75 y 100 días. Por otro lado, explorando la distribución de horas de juego se tiene un promedio de 19 horas y va de un rango de 2 a 36 horas, mientras el número de horas de entrenamiento va 12 a 186 horas.

Tabla . Indicadores de tendencia central y variación – Variables cuantitativas

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | Media | Desviación estándar | Mediana | Mínimo | Máximo | Rango |
| Edad | 16.15 | 2.46 | 16 | 12 | 20 | 8 |
| Días incapacidad | 21.63 | 23.40 | 16 | 1 | 120 | 119 |
| Talla | 171.07 | 10.70 | 171.05 | 144 | 193.5 | 49.5 |
| Peso | 65.23 | 12.11 | 67.8 | 31 | 82.5 | 51.5 |
| IMC | 22.05 | 2.28 | 22.765 | 14.95 | 25.25 | 10.3 |
| Horas juego | 19.26 | 10.54 | 22 | 2 | 36 | 34 |
| Horas entrenamiento | 96.91 | 53.31 | 110 | 12 | 186 | 174 |

Previo a la ejecución del MCA, se exploró la relación entre las variables a partir de un test de independencia Chi-cuadrado[[1]](#footnote-1) (para variables cualitativas) y coeficientes de correlación de Spearman[[2]](#footnote-2) (para variables cuantitativas). Los resultados se muestran a continuación:

|  |  |
| --- | --- |
| **A)** | **B)** |

Figura . Exploración de relaciones entre variables. A) Valor-p del test Chi-cuadrado para variables cualitativas y B) Coeficiente de correlación de Spearman para variables cuantitativas

Como se puede observar en la Figura 3-A, variables como categoría, posición, cirugía, momento de la lesión, área corporal afectada, lado, tipo de lesión, tejido afectado y gravedad muestran señales de asociación. Mientras, variables cuantitativas como talla, peso e IMC muestran una relación positiva como era de esperarse, mientras se logra distinguir levemente una relación negativa entre las horas de juego y entrenamiento con respecto a las demás variables.

Una vez, que se ha determinado la existencia de relaciones entre las variables a estudiar, el siguiente paso consistió en hacer uso del MCA para la identificación de relaciones y casos puntuales a destacar, los cuales pueden ser de interés para obtener una mejor comprensión de la naturaleza de las lesiones sobre la población estudiada. Para esto se utilizaron todas las variables cualitativas como variables activas en el análisis, mientras se utilizaron las variables cuantitativas como variables suplementarias.

En primera medida como se observa en la Figura 4, el porcentaje de variación explicado en un plano de dos dimensiones tiende a ser bajo, sin embargo, esto se debe a la presencia de variables cualitativas con un gran número de categorías lo que hace que la variación explicada tienda a dispersarse. Siendo así, los análisis posteriores cuentan con la explicación únicamente de las dos primeras dimensiones principales para dar una idea general del patrón de los datos.

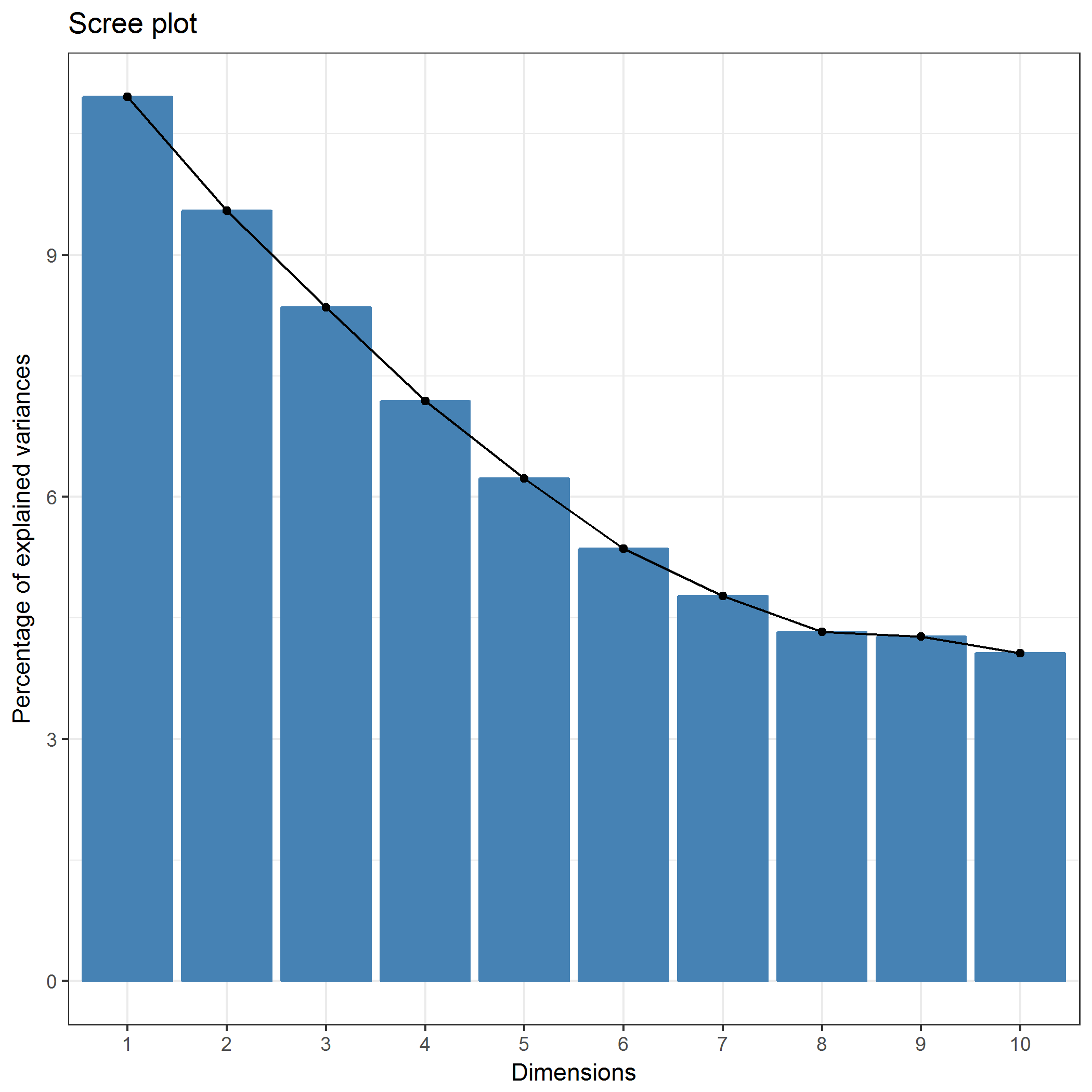


Figura . Porcentaje de variabilidad explicado por cada dimensión

Acto seguido se exploró la asociación entre categorías de variables en función de su distancia. Vale la pena aclarar que el eje X o línea horizontal hace referencia a la primera dimensión principal, mientras el eje Y o línea vertical corresponde a la segunda dimensión principal. Como se puede observar en la Figura 5, la mayor cantidad de categorías se encuentra distribuida sobre la primera dimensión principal, mientras un número más bajo se encuentran próximas a la segunda dimensión.

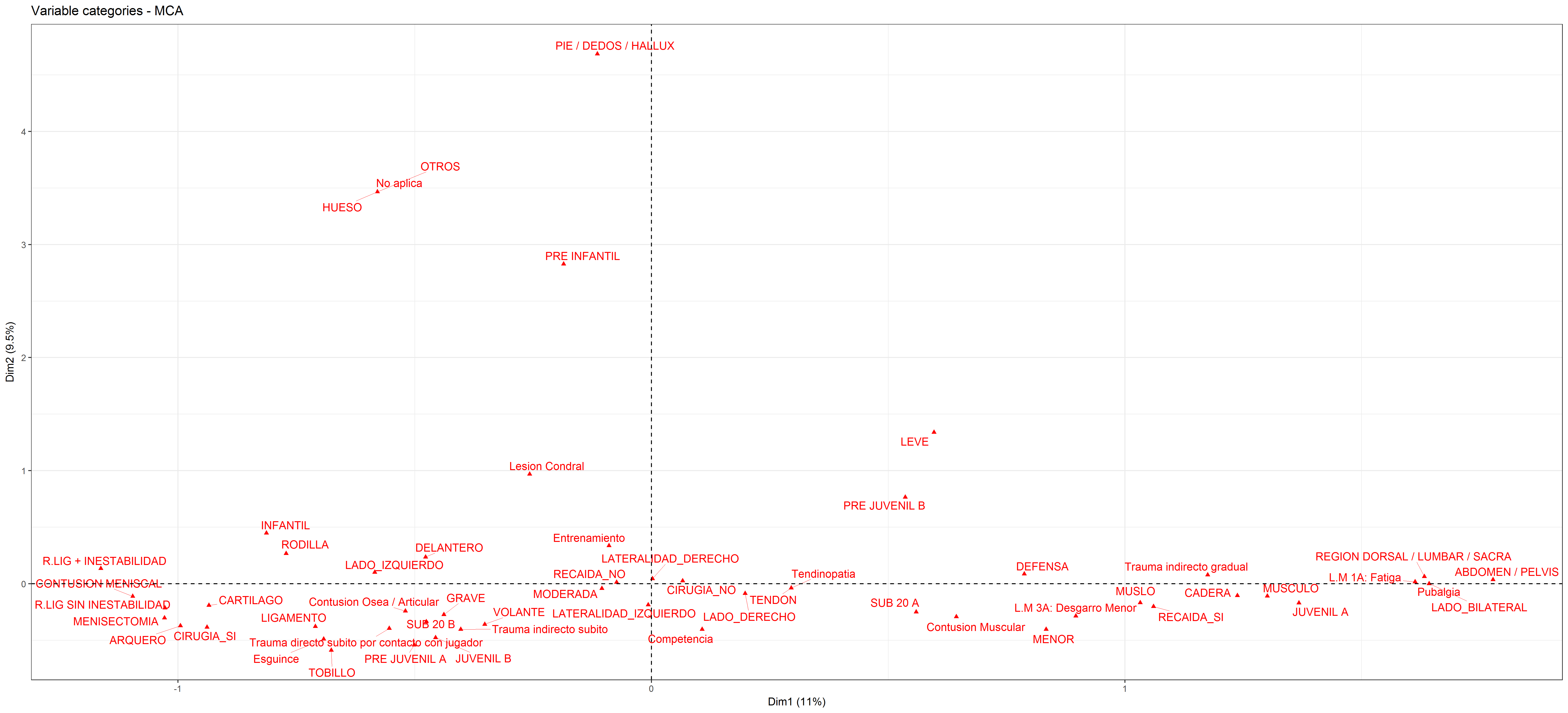


Figura . Posicionamiento variables cualitativas

En estos términos, categorías próximas indican una asociación entre ellas, como se puede observar lesiones en rodilla en delanteros y volantes tienden a presentar una gravedad que va de moderada a grave, especialmente en las categorías pre juvenil A y juvenil B. Por otro lado, los desgarros de menor gravedad tienden a presentarse en defensas y tienden a producir contusiones musculares principalmente en muslos y caderas.

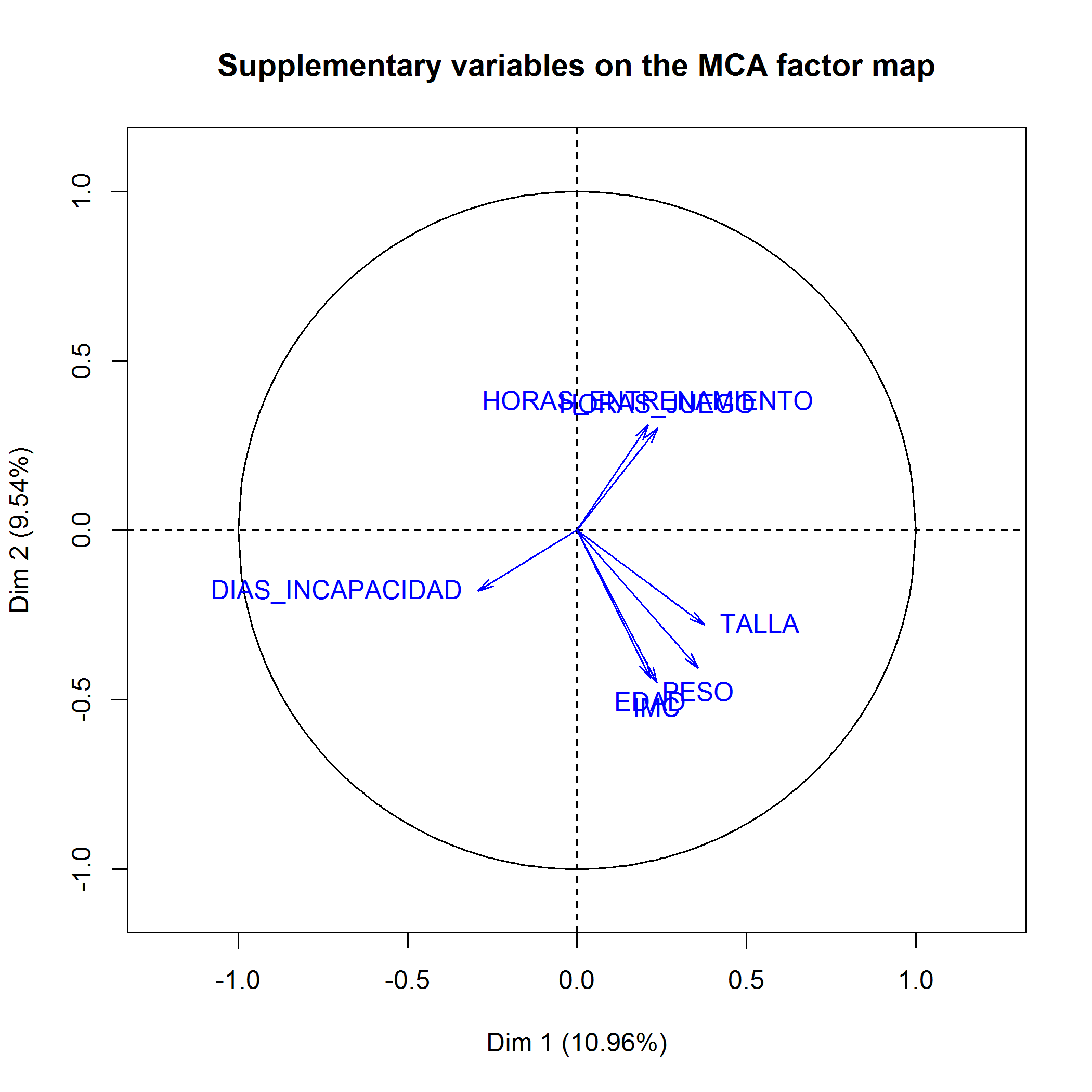


Figura . Posicionamiento variables cuantitativas

Para complementar la información previa, se explora la influencia de las variables cuantitativas en la Figura 6. De este modo, si se tiene en cuenta el posicionamiento de las categorías de la Figura 5, se tiene que los individuos que se encuentran en el cuadrante inferior derecho son los presentan una mayor edad así como peso y talla y se encuentran en categorías altas como la Sub 20. Por otro lado, los individuos situados en el cuadrante inferior izquierdo son los que presentan un mayor número de días de incapacidad y tienen asociadas lesiones más graves. Finalmente aquellos jugadores ubicados cerca del cuadrante superior derecho son los que tienen mayor número de horas de juego y entrenamiento asociados.

En último lugar, se presentan las categorías (nombres en color rojo) y los individuos (puntos de color azul) en el mismo gráfico permitiendo dar una mayor claridad sobre los casos evaluados (ver Figura 7). Un caso a destacar consiste en el jugador # 37 quien hace parte de la categoría pre infantil y cuya área corporal afectada fueron PIE / DEDOS / HALLUX. También se aclara que jugadores cercanos en el gráfico tienden a presentar un comportamiento similar, mismo tipo de lesión, posición de juego o composición física.

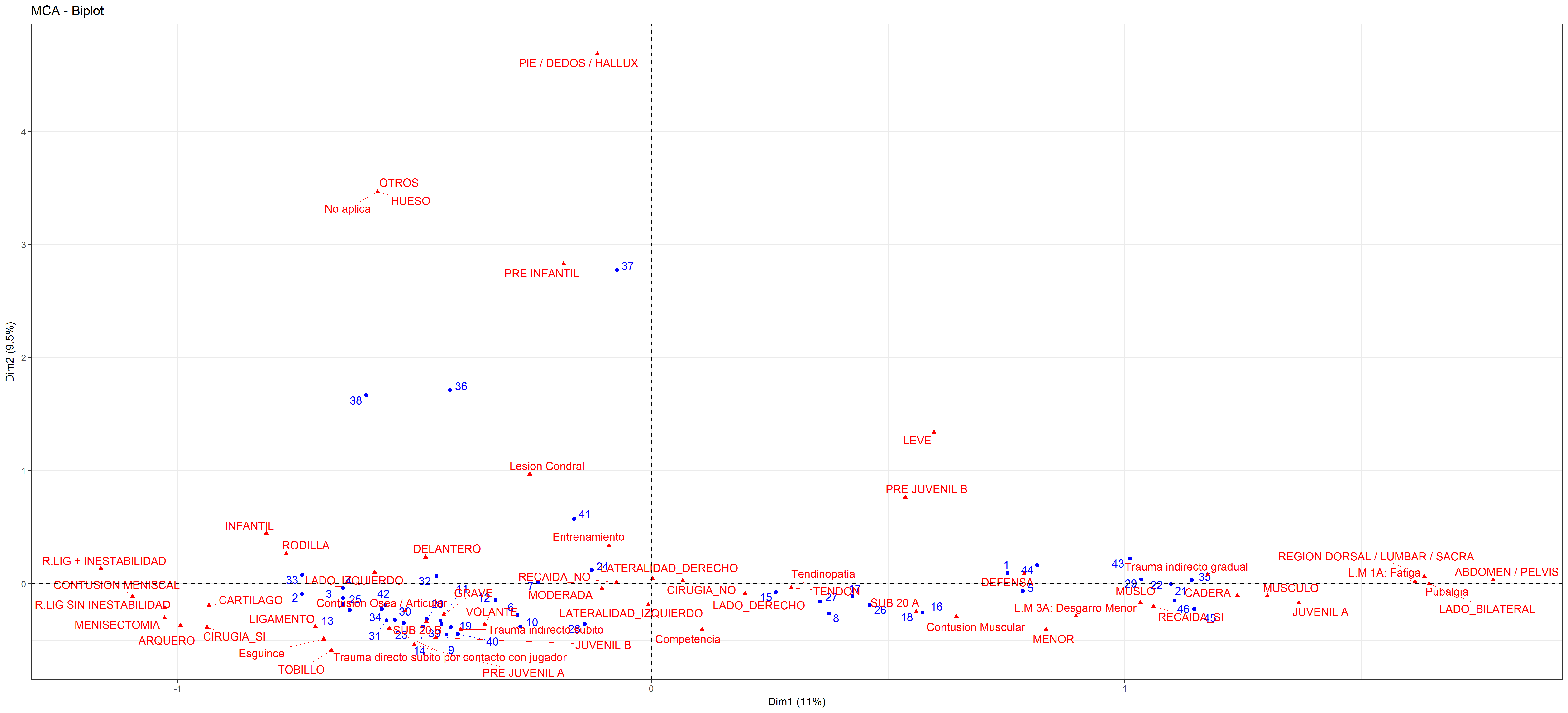


Figura . Posicionamiento conjunto individuos y variables

1. Los valores graficados corresponden al valor-p del test de independencia Chi-cuadrado, el cual mide si existe algún tipo de asociación entre variables cualitativas. De este modo, entre más próximo esté el valor-p a cero, indica que existe algún tipo de asociación entre las variables evaluadas. [↑](#footnote-ref-1)
2. El coeficiente de correlación de Spearman mide la fuerza de asociación entre dos variables cuantitativas en función de relaciones positivas o negativas. Eso es, asociaciones positivas si a medida que una variable aumenta, la otra variable en comparación lo hace del mismo modo. Y asociaciones negativas si a medida que una variable aumenta, la otra disminuye. [↑](#footnote-ref-2)