

**UNIVERSIDAD PRIVADA SAN JUAN BAUTISTA FACULTAD DE CIENCIAS  
DE LA SALUD ESCUELA PROFESIONAL DE MEDICINA HUMANA**



**ASIGNATURA:**  
SISTEMATIZACIÓN Y MÉTODOS ESTADÍSTICOS

**DOCENTE:**  
CASTRO LOPEZ, SEGUNDO VICENTE

**ESTUDIANTES:**  
CARDENAS SANCHEZ, ALAI ALEXIS  
CHOLÁN SÁNCHEZ, ARIANA MARICIELO  
PEREZ FLORES, ANDREA XIMENA  
ORTIZ ALVAREZ, MARÍA FERNANDA  
PINTO COAQUIRA, INES

**SEDE Y TURNO**  
**SAN BORJA - MA**

**2025**

## 1. ANÁLISIS DE AGRUPAMIENTO JERÁRQUICO (HIERARCHICAL CLUSTERING)

### 1.1 Dataset “Autismo”

El dataset “Autismo”, contiene información de 292 niños/as evaluados mediante un test para la detección de autismo. El dataset incluye variables numéricas, puntuaciones numéricas de 10 preguntas (ítems P1 a P10), además de variables categóricas sociodemográficas y clínicas como edad, género, etnicidad, antecedentes perinatales (Ictericia al nacer) y diagnóstico de autismo.

El objetivo es aplicar el método de agrupamiento jerárquico para identificar grupos de niños/as que compartan características similares en cuanto a su perfil de puntuaciones del test de autismo. Esto nos permitirá proponer posibles categorías de riesgo que son útiles para intervenciones más personalizadas en la detección temprana del trastorno del espectro autista (TEA).

Para ilustrar el proceso de análisis usaremos el dataset llamado “data\_autismo” el cual contiene 292 observaciones con las siguientes variables:

- Puntuaciones individuales en 10 preguntas (P1 a P10), variables numéricas binarias (0 = no presenta conducta asociada, 1 = presenta conducta asociada).
- Edad (variable numérica discreta, expresada en años; se registran valores entre 4 y 11, con algunos datos faltantes).
- Género (masculino/femenino).
- Etnicidad (variable categórica con diversas categorías como “Otros”, “Desconocido”, “Oriente Medio”, etc.).
- Ictericia al nacer (Sí/No).
- Diagnóstico previo de autismo (Sí/No).
- País de residencia (variable categórica).
- Uso previo de la aplicación (Sí/No).
- Resultado total del test (variable numérica continua del 0 al 10).
- Quién completó la prueba (por ejemplo: padre/madre, profesional de salud, etc.).
- Diagnóstico final de TEA (Sí/No).
- Descripción de la edad (variable categórica que agrupa la edad en rangos: “4-11 años”, “12-17 años”, “18 años o más”, etc.).

### 1.2 Importando los datos

```
{r}  
data_autismo <- import(here("data", "autismo.csv"))  
  
Registered S3 method overwritten by 'data.table':  
  method      from  
  print.data.table
```

### 1.3 Preparación de los datos

#### 1.3.1 Solo datos numéricos

Para el análisis de agrupamiento usaremos solo variables numéricas.

- Las puntuaciones individuales de las preguntas P1 a P10 (variables binaria: 0 o 1).
- La edad (años), una variable numérica discreta.
- El resultado total del test (de 0 a 10).

Seleccionar únicamente variables numéricas y asegurarse de que estén en el formato adecuado (por ejemplo, convertir valores faltantes o codificados como texto en Edad).

convertir la columna Edad a numérica, reemplazando los valores no válidos ("?",) por NA y asegurando que los datos estén en el formato adecuado para análisis numérico:

```
{r}
data_autismo <- data_autismo %>%
  mutate(Edad = na_if(Edad, "?"),      # Reemplaza "?" con NA
         Edad = as.numeric(Edad))     # Convierte a numérico
```

El código abajo elimina las variables categóricas como: Género, Etnicidad, Ictericia\_al\_nacer, Diagnostico\_autismo, Pais\_residencia, Uso\_prev\_aplicacion, Descripcion\_edad, Quien\_completa\_prueba y Diagnostico\_ASD

ID se conserva como identificador del participante.

```
{r}
data_autismo_1 = data_autismo |>
  select(-Género, -Etnicidad, -Ictericia_al_nacer, -Diagnostico_autismo,
         -Pais_residencia, -Uso_prev_aplicacion, -Quien_completa_prueba, -Diagnostico_ASD,
         -Descripcion_edad) |>
  column_to_rownames("ID")
```

#### 1.3.2 La importancia de estandarizar

Es importante estandarizar las variables antes de realizar el análisis de agrupamiento jerárquico.

En el caso de nuestro dataset “data\_autismo”, aunque las variables numéricas como las preguntas P1 a P10 están codificadas de forma binaria (0/1), otras variables como la edad (en años) o el resultado total del test (de 0 a 10) se encuentran en escalas diferentes, esto podría generar un sesgo en el análisis. Por ejemplo, el resultado total del test puede variar en un rango mayor que las preguntas individuales (que solo toman dos valores posibles), y si no se estandariza, esta variable podría tener mayor peso en el cálculo de distancias entre individuos.

Entonces, si se agrupa a las personas considerando simultáneamente sus respuestas a las preguntas P1-P10 (escala binaria), su edad (años) y su puntaje total del test (0-10), debemos preguntarnos: ¿Qué variable debería tener mayor influencia en la formación de los grupos?

Sin una estandarización previa, estas diferencias no serían comparables, y las variables con

mayor rango numérico podrían dominar el cálculo de distancias, afectando la validez del agrupamiento. Por ello, aplicaremos una función de estandarización, “scale()”.

- crear una matriz data autismo escalado

```
{r}
data_autismo_escalado = scale(data_autismo_1)
```

	Puntuacion_P1	Puntuacion_P2	Puntuacion_P3	Puntuacion_P4	Puntuacion_P5	Puntuacion_P6	Puntuacion_P7	Puntuacion_P8	Puntuacion_P9	Puntuacion_P10	Edad	Resultado
1	0.7592086	0.9320994	-1.6980650	-1.1067062	0.5868888	0.6343998	-1.2384908	1.0051474	-0.9847034	-1.6250922	-0.1497245	-0.5425778
2	0.7592086	0.9320994	-1.6980650	-1.1067062	0.5868888	0.6343998	-1.2384908	1.0051474	-0.9847034	-1.6250922	-0.1497245	-0.5425778
3	0.7592086	0.9320994	-1.6980650	-1.1067062	-1.6980650	0.6343998	0.8046692	1.0051474	-0.9847034	-1.6250922	-0.1497245	-0.5425778
4	-1.3126503	0.9320994	-1.6980650	-1.1067062	0.5868888	0.6343998	-1.2384908	-0.9914719	-0.9847034	0.6132423	-0.5724759	-0.9802372
5	0.7592086	0.9320994	0.5868888	0.9004877	0.5868888	0.6343998	0.8046692	1.0051474	1.0120563	0.6132423	-0.5724759	1.6457193
6	-1.3126503	-1.0691728	0.5868888	-1.1067062	0.5868888	0.6343998	-1.2384908	1.0051474	-0.9847034	0.6132423	-0.9952273	-0.5425778
7	0.7592086	-1.0691728	0.5868888	0.9004877	0.5868888	0.6343998	-1.2384908	1.0051474	-0.9847034	0.6132423	-0.5724759	0.3327411
8	0.7592086	0.9320994	0.5868888	0.9004877	0.5868888	0.6343998	0.8046692	1.0051474	-0.9847034	-1.6250922	-0.5724759	0.7704005
9	0.7592086	0.9320994	0.5868888	0.9004877	0.5868888	0.6343998	0.8046692	-0.9914719	-0.9847034	-1.6250922	1.9640327	0.3327411
10	-1.3126503	-1.0691728	0.5868888	0.9004877	0.5868888	-1.5708948	0.8046692	1.0051474	-0.9847034	-1.6250922	1.9640327	-0.5425778
11	0.7592086	-1.0691728	-1.6980650	-1.1067062	0.5868888	0.6343998	0.8046692	1.0051474	1.0120563	0.6132423	1.5412812	0.3327411
12	-1.3126503	0.9320994	-1.6980650	-1.1067062	0.5868888	-1.5708948	-1.2384908	-0.9914719	-0.9847034	0.6132423	-0.5724759	-1.4178966
13	-1.3126503	0.9320994	0.5868888	0.9004877	0.5868888	0.6343998	0.8046692	1.0051474	1.0120563	0.6132423	-0.9952273	1.2080599
14	0.7592086	-1.0691728	-1.6980650	-1.1067062	-1.6980650	-1.5708948	0.8046692	-0.9914719	-0.9847034	-1.6250922	-0.9952273	-1.8555560
15	0.7592086	0.9320994	0.5868888	0.9004877	0.5868888	0.6343998	0.8046692	1.0051474	1.0120563	0.6132423	-0.1497245	1.6457193
16	0.7592086	0.9320994	0.5868888	0.9004877	0.5868888	0.6343998	0.8046692	1.0051474	1.0120563	0.6132423	0.6957784	1.6457193
17	0.7592086	0.9320994	0.5868888	0.9004877	0.5868888	0.6343998	-1.2384908	1.0051474	1.0120563	0.6132423	-0.9952273	1.2080599
18	-1.3126503	-1.0691728	-1.6980650	-1.1067062	-1.6980650	-1.5708948	0.8046692	-0.9914719	-0.9847034	-1.6250922	0.2730270	-2.2932154
19	0.7592086	-1.0691728	0.5868888	0.9004877	0.5868888	-1.5708948	0.8046692	1.0051474	1.0120563	0.6132423	1.9640327	0.7704005

## INTERPRETACIÓN:

La tabla nos muestra un conjunto de datos llamado data\_autismo\_escalado, que representa valores estandarizados (escalados) de las variables originales. La estandarización (usando la función “scale()”) transforma los datos para que tengan media 0 y desviación estándar 1.

**Filas:** 292 observaciones (casos/individuos)

**Columnas:** 12 variables (11 predictores + 1 resultado)

**Variables:**

**Dimensiones\_P1 a P10:** Probablemente son ítems de un cuestionario o evaluación relacionada con el autismo (10 dimensiones)

**Edad:** Variable demográfica escalada

**Resultado:** Variable de resultado/diagnóstico escalada

### Observaciones

- **Valores escalados:** Todos los números representan desviaciones estándar de la media. Por ejemplo:
  - Un valor de 1.54 en "Edad" (fila 10) indica que esa persona tiene una edad 1.54 desviaciones estándar por encima de la media
  - Un valor de -1.31 en "Dimensions\_P1" (varias filas) está 1.31 desviaciones por debajo de la media
- **Patrones interesantes:**
  - Las filas 1-3 son casi idénticas excepto en Dimensions\_P7
  - La fila 18 tiene valores extremos en varias dimensiones (P5, P6).
  - Los resultados varían desde -1.47 (fila 11) hasta 2.29 (fila 18)

- **Relaciones potenciales:**

- Los casos con valores positivos en múltiples dimensiones (ej. filas 5, 14-17) tienden a tener resultados positivos
- Los casos con valores negativos en varias dimensiones (ej. filas 4, 6, 11) tienden a tener resultados negativos

- Un vistazo a los datos antes del escalamiento:

```
{r}
head(data_autismo_1)
```

Description: df [6 × 12]

	Puntuacion_P8 <int>	Puntuacion_P9 <int>	Puntuacion_P10 <int>	Edad <dbl>	Resultado <int>
1	1	0	0	6	5
2	1	0	0	6	5
3	1	0	0	6	5
4	0	0	1	5	4
5	1	1	1	5	10
6	1	0	1	4	5

6 rows | 9-13 of 12 columns

- Un vistazo a los datos después del escalamiento:

```
{r}
head(data_autismo_escalado)
```

	Puntuacion_P1	Puntuacion_P2	Puntuacion_P3	Puntuacion_P4	Puntuacion_P5	Puntuacion_P6	Puntuacion_P7
1	0.7592086	0.9320994	-1.6980650	-1.1067062	0.5868888	0.6343998	-1.2384908
2	0.7592086	0.9320994	-1.6980650	-1.1067062	0.5868888	0.6343998	-1.2384908
3	0.7592086	0.9320994	-1.6980650	-1.1067062	-1.6980650	0.6343998	0.8046692
4	-1.3126503	0.9320994	-1.6980650	-1.1067062	0.5868888	0.6343998	-1.2384908
5	0.7592086	0.9320994	0.5868888	0.9004877	0.5868888	0.6343998	0.8046692
6	-1.3126503	-1.0691728	0.5868888	-1.1067062	0.5868888	0.6343998	-1.2384908

	Puntuacion_P8	Puntuacion_P9	Puntuacion_P10	Edad	Resultado
1	1.0051474	-0.9847034	-1.6250922	-0.1497245	-0.5425778
2	1.0051474	-0.9847034	-1.6250922	-0.1497245	-0.5425778
3	1.0051474	-0.9847034	-1.6250922	-0.1497245	-0.5425778
4	-0.9914719	-0.9847034	0.6132423	-0.5724759	-0.9802372
5	1.0051474	1.0120563	0.6132423	-0.5724759	1.6457193
6	1.0051474	-0.9847034	0.6132423	-0.9952273	-0.5425778

## INTERPRETACIÓN:

Podemos observar el proceso de preprocesamiento de datos, específicamente el escalamiento. El conjunto de datos original (`data_autismo_1`) contenía características con diferentes rangos y tipos de valores. Después del escalamiento, el conjunto de datos (`data_autismo_escalado`) ha sido transformado para que todas las características tengan una escala comparable, lo que es un paso esencial antes de aplicar la mayoría de los modelos de aprendizaje automático para asegurar un rendimiento óptimo y justo. La presencia de valores negativos nos indica que se usó la estandarización.

### 1.4 Cálculo de distancias

Se calcula la distancia entre cada par de individuos del dataset en función de sus características numéricas estandarizadas (por ejemplo, sus respuestas al test, edad y puntaje total).

- Función “`dist()`” para calcular la distancia euclidiana entre cada par de observaciones. El resultado de este cálculo es una matriz de distancias o disimilitud, que representa qué tan “lejanos” o “ceranos” están unos participantes de otros, en el espacio definido por las variables numéricas.

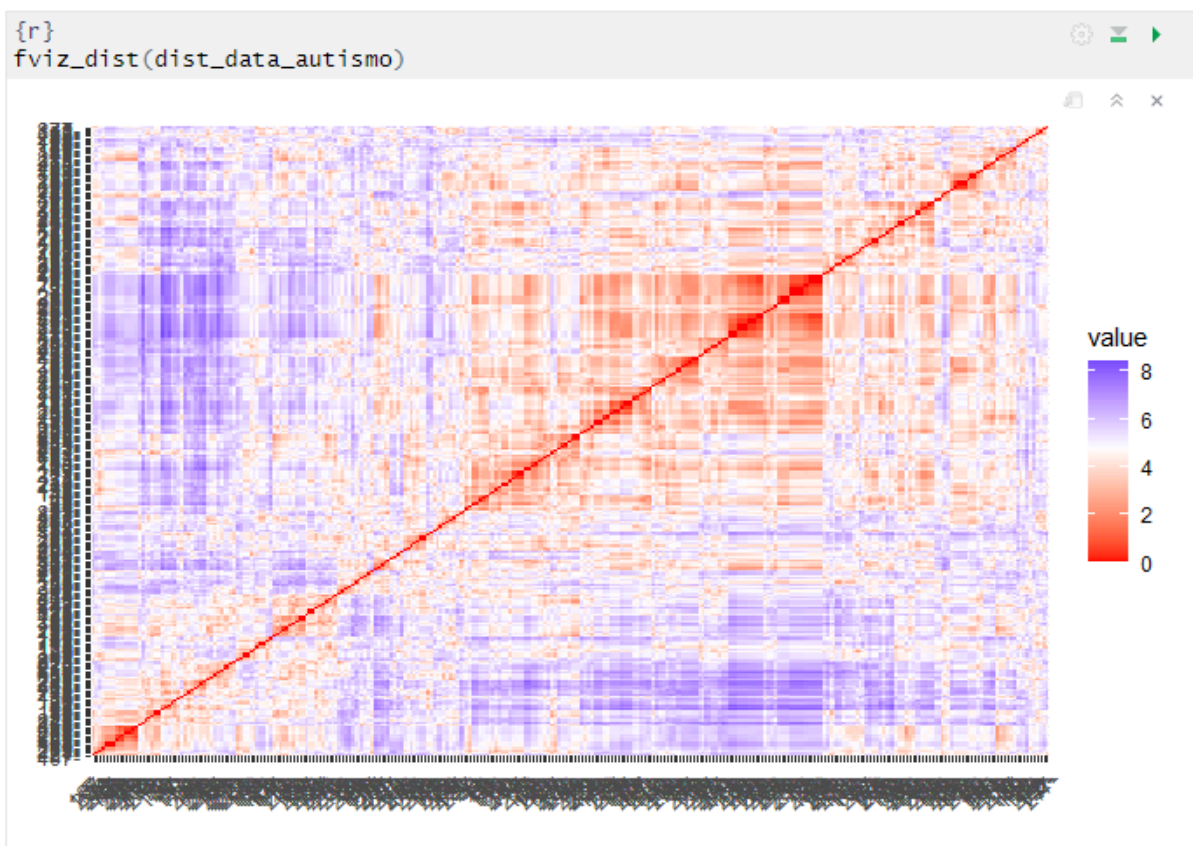
```
{r}  
dist_data_autismo <- dist(data_autismo_escalado, method = "euclidean")
```

values	
dist_data_autismo	'dist' num [1:42486] 0 3.07 3.7 5.23 4.39 ...

#### 1.4.1 Visualizando las distancias euclidianas con un mapa de calor

Usamos mapas de calor (heatmaps) para visualizar si existen patrones de agrupamiento

- Función “`fviz_dist()`” del paquete `factoextra` para crear un mapa de calor.



### INTERPRETACIÓN:

Este mapa de calor representa la matriz de distancias euclidianas entre todos los participantes del dataset de autismo, basándose en sus características numéricas estandarizadas (como la edad, el puntaje total del test y las respuestas P1 a P10). Podemos observar el mapa de calor de la matriz de distancias para `dist_data_autismo`, nos revela la estructura de similitud/disimilitud dentro del conjunto de datos. La presencia de regiones con distancias bajas (rojo/naranja) nos indica que existen grupos o clústeres de individuos con autismo que comparten características muy similares, mientras que las regiones con distancias más altas (púrpura) indican diferencias significativas entre otros individuos o grupos. Esta visualización es un paso exploratorio común para entender la estructura intrínseca de los datos antes de aplicar técnicas de agrupamiento o clasificación.

Este mapa de calor representa la matriz de distancias euclidianas entre todos los participantes del dataset de autismo, basándose en sus características numéricas estandarizadas (como la edad, el puntaje total del test y las respuestas P1 a P10).

- Rojo/naranja (cercano a 0): indica alta similitud entre dos participantes. Es decir, tienen perfiles muy parecidos en las variables consideradas.

- Púrpura (valores altos): indica alta disimilitud, es decir, los participantes son muy diferentes entre sí.
- Blanco o colores intermedios: representan distancias moderadas.

**Diagonal roja:** La línea diagonal roja que cruza el gráfico de arriba a la derecha representa las comparaciones de cada individuo consigo mismo, por lo tanto, todas las distancias son 0 (máxima similitud).

**Bloques claros o rojizos cerca de la diagonal:** Se observan zonas más claras o rojizas agrupadas, nos indica que existen subconjuntos de individuos con perfiles similares.

**Áreas azuladas/púrpuras dispersas:** Estas indican pares de individuos muy diferentes entre sí. Cuando se observa una zona amplia con mucho azul, eso implica mayor variabilidad entre los perfiles.

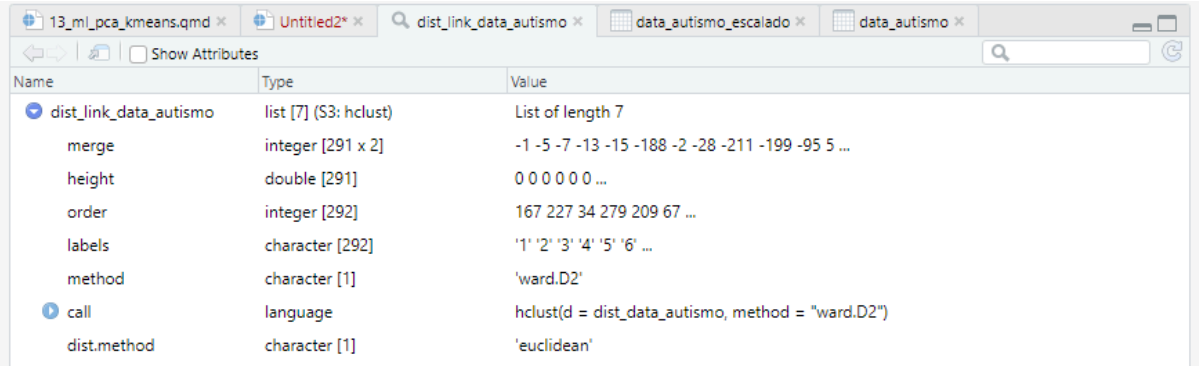
### 1.5. Método de agrupamiento: función de enlace (linkage)

Técnica que comienza uniendo las observaciones más similares entre sí, lo que lo hace intuitivo y visualmente interpretativo. Una vez que se forman los primeros grupos (clústeres), debemos decidir cómo medir la distancia entre estos nuevos grupos.

- Función de enlace (linkage), utiliza la matriz de distancias generada previamente para determinar cómo agrupar sucesivamente los participantes o grupos, en función de su similitud.
- Métodos para realizar agrupamiento: Enlace máximo o completo, Enlace mínimo o simple, Enlace de la media o promedio, Enlace de centroide, Método de varianza mínima de Ward

En este análisis de agrupamiento jerárquico con datos de las pruebas de autismo, usaremos el método de varianza mínima de Ward

```
{r}
dist_link_data_autismo <- hclust(d = dist_data_autismo, method = "ward.D2")
```

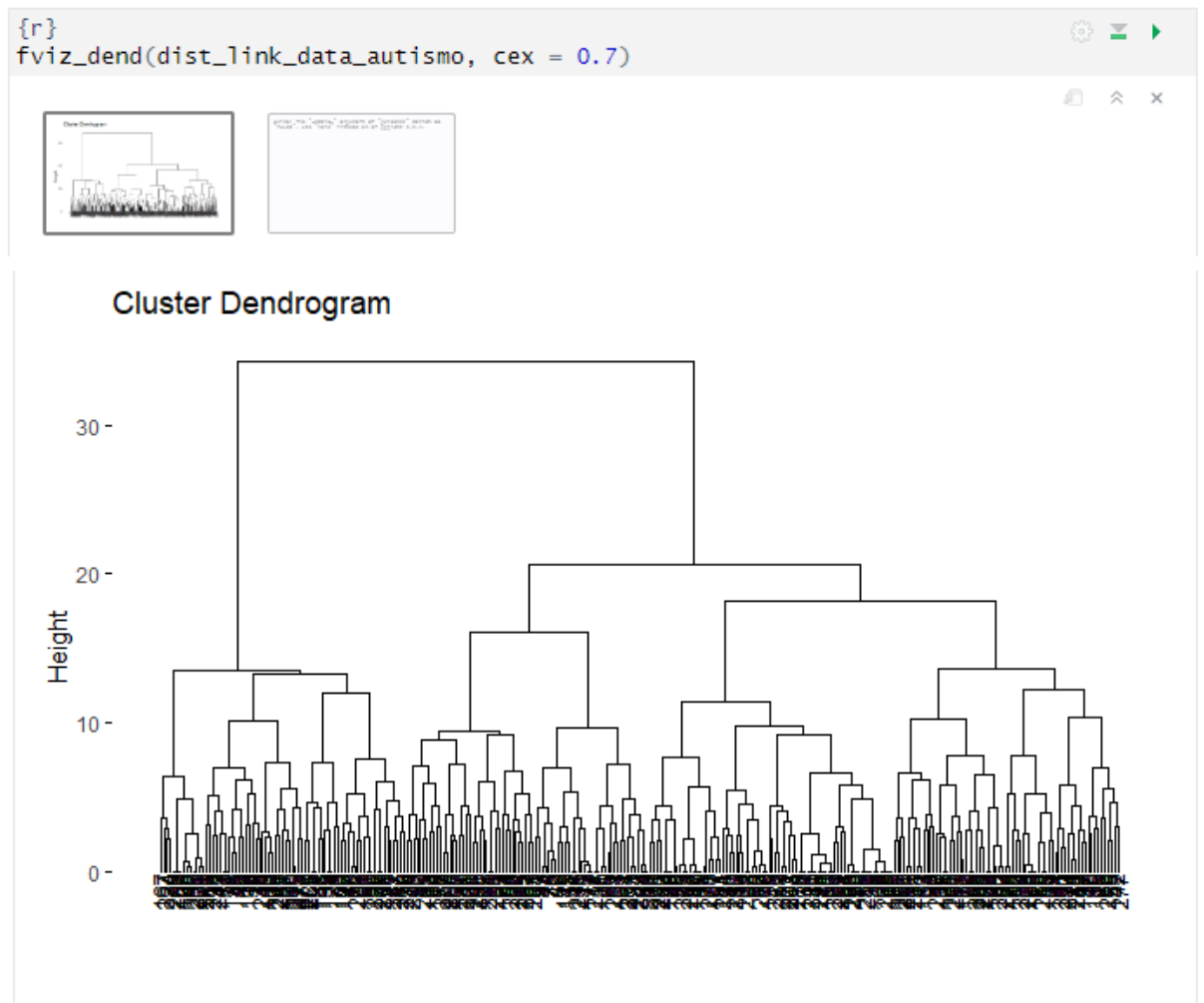


Name	Type	Value
dist_link_data_autismo	list [7] (S3: hclust)	List of length 7
merge	integer [291 x 2]	-1 -5 -7 -13 -15 -188 -2 -28 -211 -199 -95 5 ...
height	double [291]	0 0 0 0 0 ...
order	integer [292]	167 227 34 279 209 67 ...
labels	character [292]	'1' '2' '3' '4' '5' '6' ...
method	character [1]	'ward.D2'
call	language	hclust(d = dist_data_autismo, method = "ward.D2")
dist.method	character [1]	'euclidean'



## 1.6 Dendrogramas para la visualización de patrones

- Función `hclust()`. Para generar la representación gráfica del árbol jerárquico



### INTERPRETACIÓN:

Este análisis nos demuestra un enfoque metódico para identificar grupos homogéneos dentro del conjunto de datos de autismo. Al estandarizar los datos, visualizar sus similitudes y aplicar un método de clustering jerárquico, se busca descubrir patrones subyacentes que podrían ser cruciales para comprender la heterogeneidad del trastorno del espectro autista y, potencialmente, informar estrategias de investigación o intervención más específicas.

El dendrograma mostrado representa la estructura jerárquica del agrupamiento de los participantes evaluados con el test de prueba para autismo, basada en sus respuestas numéricas estandarizadas.

- Cada línea en la parte inferior representa a un individuo (participante).
- A medida que subimos en el gráfico, las ramas se fusionan en grupos más grandes basándose en la similitud de sus respuestas.

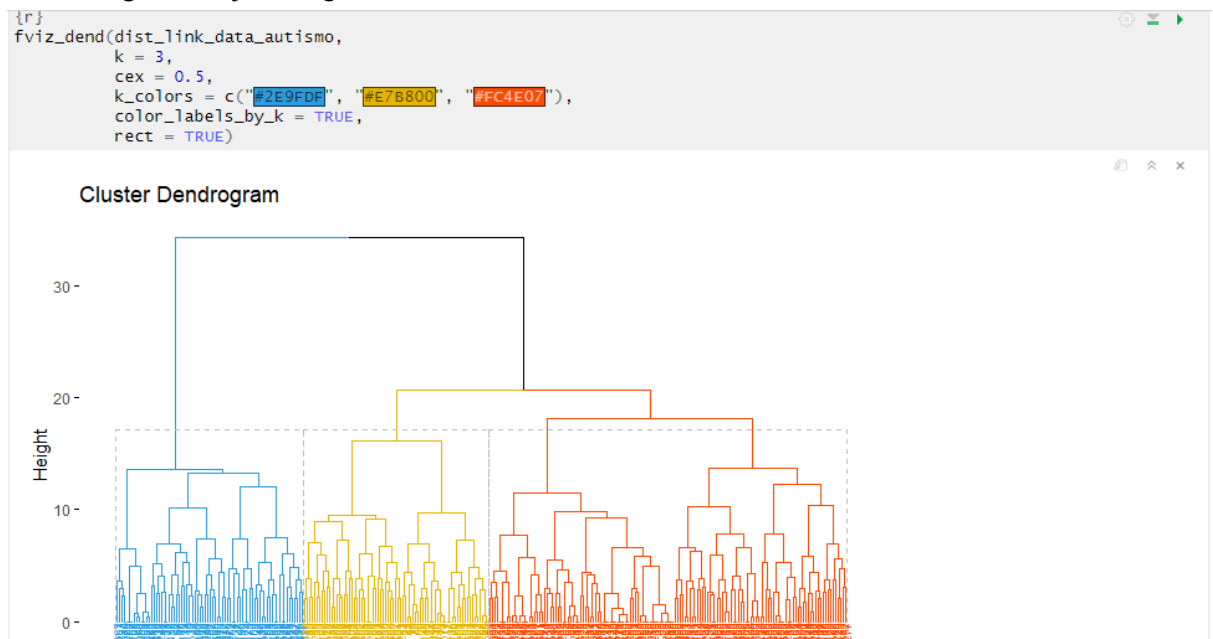
- El eje vertical ("Height") representa la distancia o disimilitud entre los clústeres que se van uniendo. Una altura mayor indica que los grupos fusionados en ese punto son más distintos entre sí.

Este dendrograma nos indica que dentro del grupo evaluado hay estructuras internas diferenciadas

### 1.7 ¿Cuántos grupos se formaron en el dendrograma?

Para nuestro dendrograma, el dendrograma muestra tres grupos.

En el código de abajo, el argumento  $k = 3$  define el número de clusters.



### INTERPRETACIÓN:

#### Número de grupos:

- Se identificaron claramente 3 agrupaciones distintas en la estructura jerárquica del dendrograma.
- Esto fue confirmado por el parámetro  $k = 3$  en el código utilizado, que específicamente solicita esta cantidad de clusters.

### 2. Implicaciones del dendrograma:

- La formación de 3 grupos sugiere que los datos tienen una estructura natural que se divide en tres categorías distintas.
- Cada grupo representa observaciones con características similares según las variables analizadas.

### 3. Sobre el parámetro $k=3$ :

- Este valor no es arbitrario, normalmente se selecciona basado en:
- El método del codo
- Análisis visual del dendrograma
- Criterios de validación de clusters
- Requerimientos específicos del análisis

### 4. Posibles interpretaciones:

- Podría representar tres niveles distintos de características relacionadas con autismo (leve, moderado, severo)
- O tres subtipos diferentes dentro del espectro autista
- O grupos basados en patrones de respuesta a las dimensiones evaluadas

## 2. AGRUPAMIENTO CON EL ALGORITMO K-MEANS (DATASET AUTISMO)

El número de grupos debe definirse de antemano. Se ha decidido formar tres grupos ( $K = 3$ ), en base a la estructura observada previamente en el dendrograma. Se aplicó el algoritmo de K-means al conjunto de datos estandarizado, que contiene las respuestas a las pruebas de detección de autismo. El objetivo es identificar patrones de respuesta similares entre los participantes, que podrían corresponder a distintos perfiles clínicos o conductuales relacionados con el espectro autista.

### 2.1 Estimando el número óptimo de clusters

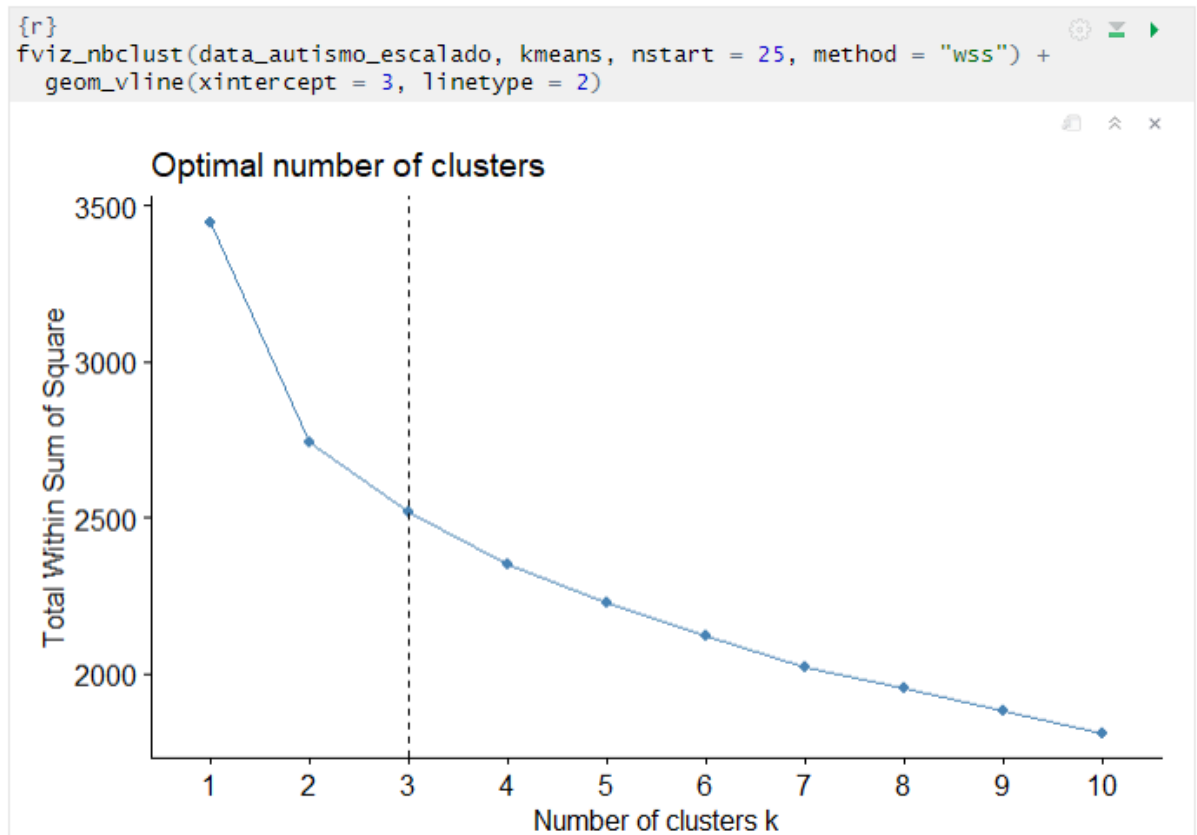
Agrupamiento k-means probando diferentes valores de  $k$  (número de clústeres). Luego, se grafica la suma de cuadrados dentro de los clústeres (WSS) en función del número de clústeres.

Función “fviz\_nbclust()”: para estimar el número óptimo de clústeres.

- Primero escalamos los datos:

```
{r}  
data_autismo_escalado = scale(data_autismo_1)
```

- Ahora graficamos la suma de cuadrados dentro de los gráficos:



**INTERPRETACIÓN:** El punto donde la curva forma una "rodilla" o quiebre, indica el número óptimo de clústeres. Para nuestro gráfico, es en el número de cluster 3.

## 2.2 Cálculo del agrupamiento k-means

Dado que el resultado final del agrupamiento k-means es sensible a las asignaciones aleatorias iniciales, se especifica el argumento `nstart = 25`. Esto significa que R intentará 25 asignaciones aleatorias diferentes y seleccionará la mejor solución, es decir, aquella con la menor variación dentro de los clústeres. El valor predeterminado de `nstart` en R es 1. Sin embargo, se recomienda ampliamente utilizar un valor alto, como 25 o 50, para obtener un resultado más estable y confiable.

```
{r}
set.seed(123)
km_res <- kmeans(data_autismo_escalado, 3, nstart = 25)
```



```

within cluster sum of squares by cluster:
[1] 729.4714 824.0789 961.9642
(between_SS / total_SS = 27.0 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
"betweenss"    "size"         "iter"
[9] "ifault"

```

### INTERPRETACIÓN:

Se realizó un análisis de clúster utilizando el algoritmo k-means con los datos escalados, agrupando a los individuos en tres clústeres. Los tamaños de los clústeres resultaron ser 110, 80 y 98 personas, respectivamente. Los centroides de cada clúster mostraron diferencias en las puntuaciones promedio de las variables evaluadas.

**El resultado muestra dos cosas:**

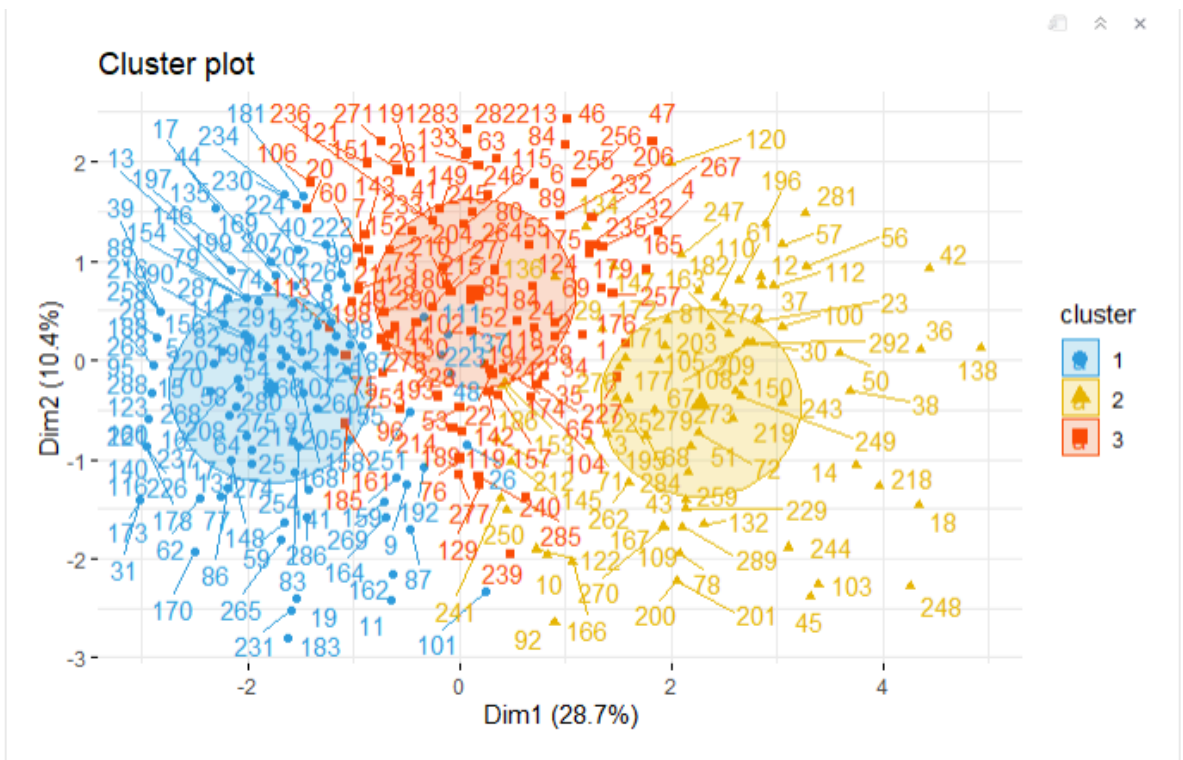
1. **Las medias o centros de los clústeres** (*Cluster means*): una matriz cuyas filas corresponden al número de clúster (1 a 3) y cuyas columnas representan las variables.
2. **Un vector de asignación de clúster** (*Clustering vector*): un vector de números enteros (de 1 a 3) que indica a qué clúster ha sido asignado cada punto (para nuestro dataset, cada paciente).

### 2.3 Visualización de los clústeres k-means

Los datos se pueden representar en un gráfico de dispersión, coloreando cada observación según el clúster al que pertenece.

- Función “fviz\_cluster()” del paquete factoextra: para visualizar los clusters generados por k-means. Esta función toma como argumentos los resultados del k-means y los datos originales (data\_autismo\_escalado).

```
{r}
fviz_cluster(
  km_res,
  data = data_autismo_escalado,
  palette = c("#2E9FDF", "#E7B800", "#FC4E07"),
  ellipse.type = "euclid",
  repel = TRUE,
  ggtheme = theme_minimal()
)
```



## INTERPRETACIÓN:

Este grafico es una visualización bidimensional de los resultados de un análisis de clúster (K-means) aplicado al conjunto de datos `data_autismo_escalado`. Los ejes Dim1 y Dim2 provienen de una técnica de reducción de dimensionalidad (como PCA o MDS) y facilitan la interpretación visual de los grupos detectados.

### Interpretación de cada clúster

- **Clúster 1 (color azul):** Ubicado hacia la izquierda en el eje X. Presenta una dispersión moderada representada por una elipse. Indica un grupo con perfiles similares, bastante homogéneo en términos de las variables originales.
- **Clúster 2 (color amarillo)**  
Se encuentra hacia la derecha, con valores positivos en Dim1. Es un grupo extenso y más disperso, lo que sugiere una mayor variabilidad interna. Representa un subgrupo diverso.
- **Clúster 3 (color rojo)**  
Ocupa la zona central del gráfico, actuando de manera intermedia entre los otros dos. Su dispersión es moderada y podría corresponder a perfiles mixtos o de transición entre los extremos.

proviene de una técnica de reducción de dimensionalidad (como PCA o MDS) y facilitan la interpretación visual de los grupos detectados.

**Ejes:**

- **Dim1 (eje horizontal):**explica aproximadamente el 28,7 % de la varianza total.
- **Dim2 (eje vertical):**explica cerca del 10,4 %.

Nos permiten distinguir los grupos con diferencias estructurales significativas en el espacio reducido.

- La separación principal ocurre a lo largo de Dim1, con Clúster 1 y Clúster 2
- Clúster 3 actúa como puente o punto intermedio, capturando perfiles que comparten características con los otros dos grupos.
- La dispersión de cada clúster indica su grado de homogeneidad o heterogeneidad interna. El Clúster 1 es el más compacto, el 2 el más amplio, y el 3 tiene una distribución intermedia.