

## 2.3.MEDIDAS DE DIAGNÓSTICO

El objetivo de las medidas de diagnóstico es identificar subconjuntos de datos que parezcan tener una influencia desproporcionada en la estimación del modelo y establecer que partes del modelo estimado se ven más afectadas por estos subconjuntos. Se debe estudiar el que un pequeño subconjunto de datos pueda tener una influencia desproporcionada en la estimación de los parámetros o en las predicciones ya que, si este es el caso, es bastante probable que los modelos estimados estén basados principalmente en estos subconjuntos en lugar de en la mayoría de los datos.

La matriz Hat queda definida para cada individuo de la siguiente forma:

$$\mathbf{H}_i = \mathbf{W}_i^{\frac{1}{2}} \mathbf{X}_i \mathbf{J}_1 \mathbf{X}_i^T \mathbf{W}_i^{\frac{1}{2}}$$

con

$$\mathbf{W}_i = \mathbf{\Lambda}_i \mathbf{V}_i^{-1} \mathbf{\Lambda}_i, \quad \mathbf{\Lambda}_i = \text{diag} \left( \frac{d\mu_i}{d\eta_i^T} \right)$$

y

$$\mathbf{J}_1 = \left[ \sum_{i=1}^n \left( \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} \right)^T \hat{\mathbf{V}}_i^{-1} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} \right]^{-1}$$

Un punto leverage es una observación grande en comparación con las otras observaciones. El “leverage” del individuo  $i$  está contenido en la matriz  $\mathbf{H}_i$  definida anteriormente. Una medida del “leverage” del individuo  $i$  puede ser definida como la traza de  $\mathbf{H}_i$

$$\text{Leverage} C_i = \text{traza}(\mathbf{H}_i)$$

La influencia del individuo  $i$  en los coeficientes de regresión puede ser valorada con la distancia de Cook

$$\text{Cook} C_i = \left( \boldsymbol{\beta}_{EEG} - \hat{\boldsymbol{\beta}}_{EEG}^{(i)} \right)^T \mathbf{J}_1^{-1} \left( \boldsymbol{\beta}_{EEG} - \hat{\boldsymbol{\beta}}_{EEG}^{(i)} \right) \frac{1}{p\hat{\gamma}}$$

Donde  $\hat{\boldsymbol{\beta}}_{EEG}^{(i)}$  denota el parámetro de regresión estimado cuando el individuo  $i$  es eliminado de los datos y  $p$  es el número de parámetros de regresión. Esta medida se puede aproximar por

$$\begin{aligned} \text{Cook} C_i &= (\mathbf{Y}_i - \boldsymbol{\mu}_i)^T \mathbf{\Lambda}_i^{-1} \left( \mathbf{W}_i^{-1} - \mathbf{X}_i \mathbf{J}_1^{-1} \mathbf{X}_i^T \right)^{-1} \mathbf{X}_i \mathbf{J}_1^{-1} \mathbf{X}_i^T \\ &\quad \times \left( \mathbf{W}_i^{-1} - \mathbf{X}_i \mathbf{J}_1^{-1} \mathbf{X}_i^T \right)^{-1} \mathbf{\Lambda}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) \frac{1}{p\hat{\gamma}} \end{aligned}$$

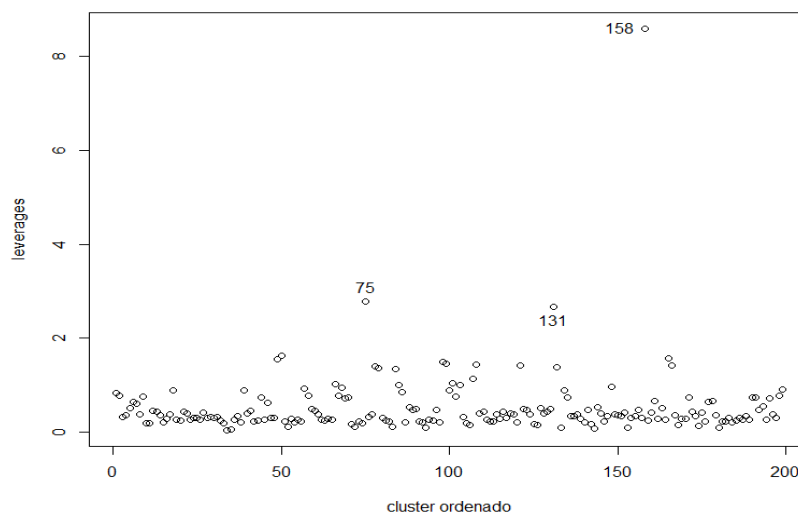
La medida de influencia DFFITC<sub>*i*</sub> de una observación es una medida de la influencia del individuo  $i$  sobre el ajuste global. Este nombre lo dieron Besley et al. [1] (difference between fitted values, es decir, distancia entre valores predichos).

$$\text{DFFIT} C_i = (\mathbf{Y}_i - \boldsymbol{\mu}_i)^T \mathbf{\Lambda}_i^{-1} \left( \mathbf{W}_i^{-1} - \mathbf{X}_i \mathbf{J}_1^{-1} \mathbf{X}_i^T \right)^{-1} \mathbf{H}_i \mathbf{\Lambda}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) \frac{1}{p\hat{\gamma}}$$

Con el fin de estudiar cómo funcionan las medidas de diagnóstico en las EEG, hemos programado algunas de las más conocidas para diferentes estructuras de la matriz de correlación de trabajo, y hemos comparado los resultados con los obtenidos directamente con el MLG, es decir, con matriz de correlación de trabajo independiente. Los resultados obtenidos se muestran a continuación:

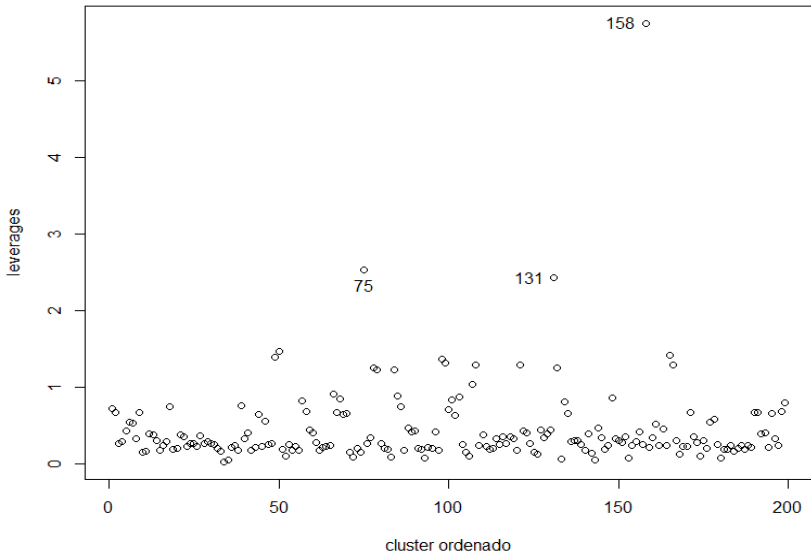
La primera de las gráficas corresponde a la medida LeverageCi con matriz de correlación de trabajo AR1. Según esta medida la observación número 158 es un punto leverage. Se podrían considerar leverages también los individuos 75 y 131.

**GRÁFICA 1. MEDIDA LEVERAGECI PARA NEPAL CON MATRIZ DE CORRELACIÓN DE TRABAJO AR1**



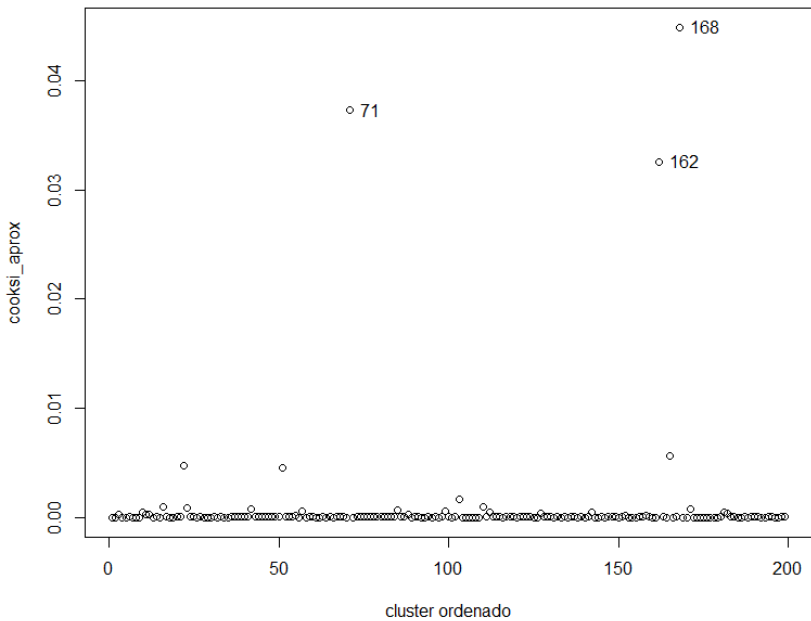
La segunda de las gráficas corresponde a la medida LeverageCi con matriz de correlación de trabajo independiente. Según esta medida la observación número 158 es un punto leverage. Se podrían considerar leverages también los individuos 75 y 131.

**GRÁFICA 5. MEDIDA LEVERAGECI PARA NEPAL CON MATRIZ DE CORRELACIÓN DE TRABAJO INDEPENDIENTE**



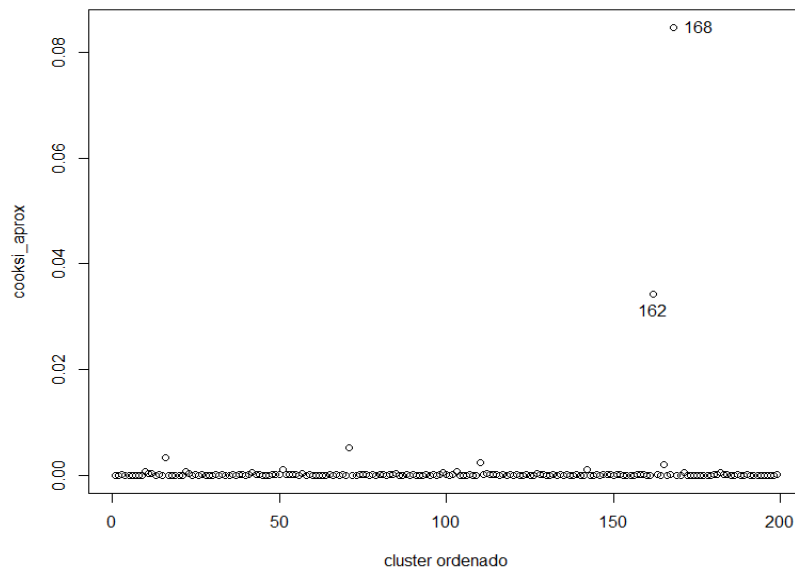
La tercera de las gráficas corresponde a la distancia de Cook con matriz de correlación de trabajo AR1. Según esta medida hay tres individuos que influyen más que el resto en los coeficientes de regresión: 71, 162 y 168.

**GRÁFICA 6. DISTANCIA DE COOK PARA NEPAL CON MATRIZ DE CORRELACIÓN DE TRABAJO AR1**



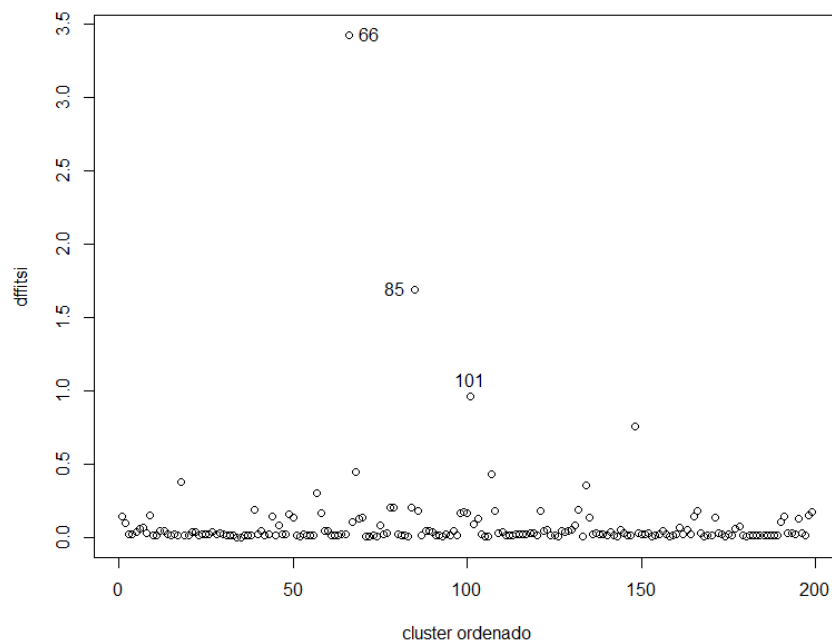
La cuarta de las gráficas corresponde a la distancia de Cook con matriz de correlación de trabajo independiente. Según esta medida hay tres individuos que influyen más que el resto en los coeficientes de regresión: 162 y 168.

**GRÁFICA 7. DISTANCIA DE COOK PARA NEPAL CON MATRIZ DE CORRELACIÓN DE TRABAJO INDEPENDIENTE**



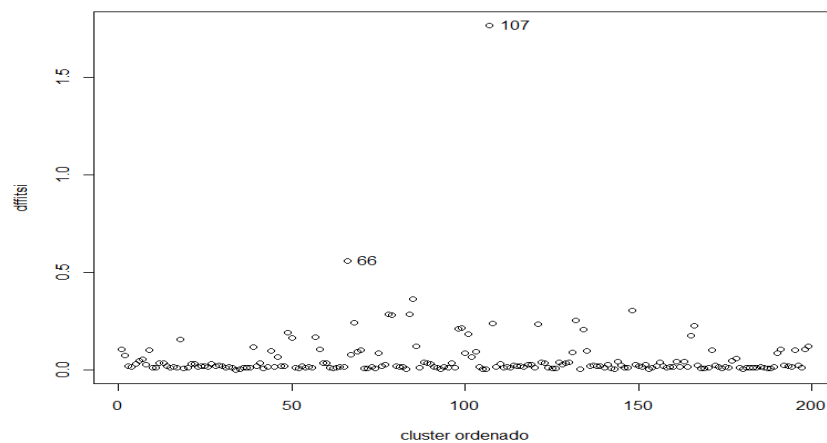
La quinta de las gráficas corresponde a la medida de influencia DFFITCi con matriz de correlación de trabajo AR1. Según esta medida hay tres observaciones que influyen sobre el ajuste global: 66, 85 y 101.

**GRÁFICA 8. MEDIDA DE INFLUENCIA DFFITCi PARA NEPAL CON MATRIZ DE CORRELACIÓN DE TRABAJO AR1**



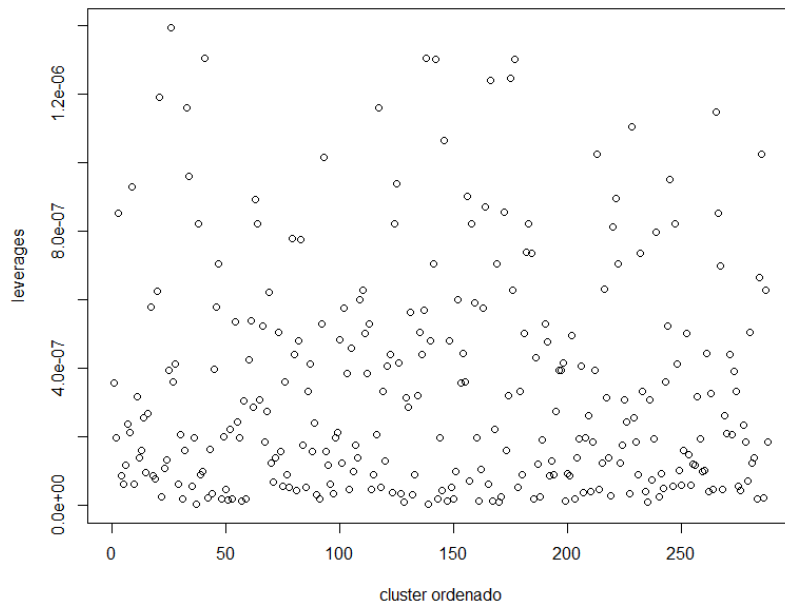
Por último, la quinta de las gráficas corresponde a la medida de influencia  $DFFITCi$  con matriz de correlación de trabajo independiente. Según esta medida hay tres observaciones que influyen sobre el ajuste global: 66 y 107.

**GRÁFICA 9. MEDIDA DE INFLUENCIA DFFITCi PARA NEPAL CON MATRIZ DE CORRELACIÓN DE TRABAJO INDEPENDIENTE**



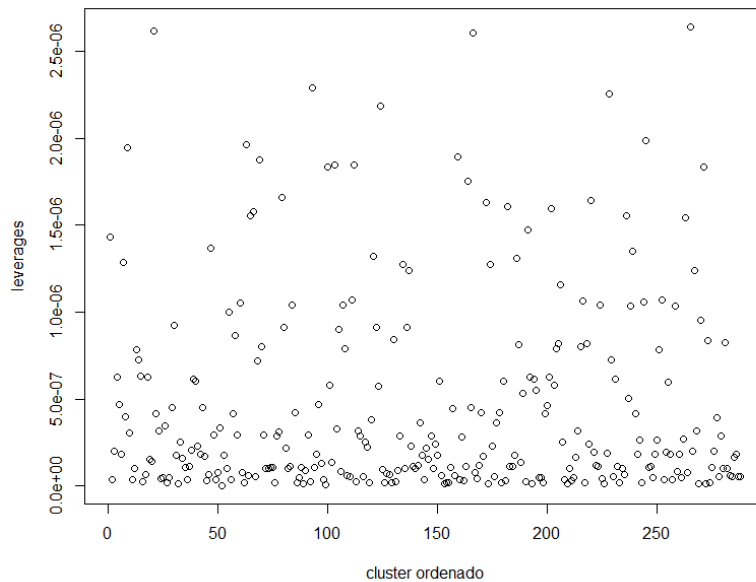
La primera de las gráficas corresponde a la medida  $LeverageCi$  con matriz de correlación de trabajo AR1. Según esta medida no hay puntos leverages.

**GRÁFICA 10. MEDIDA LEVERAGE PARA SMOKE CON MATRIZ DE CORRELACIÓN DE TRABAJO AR1**



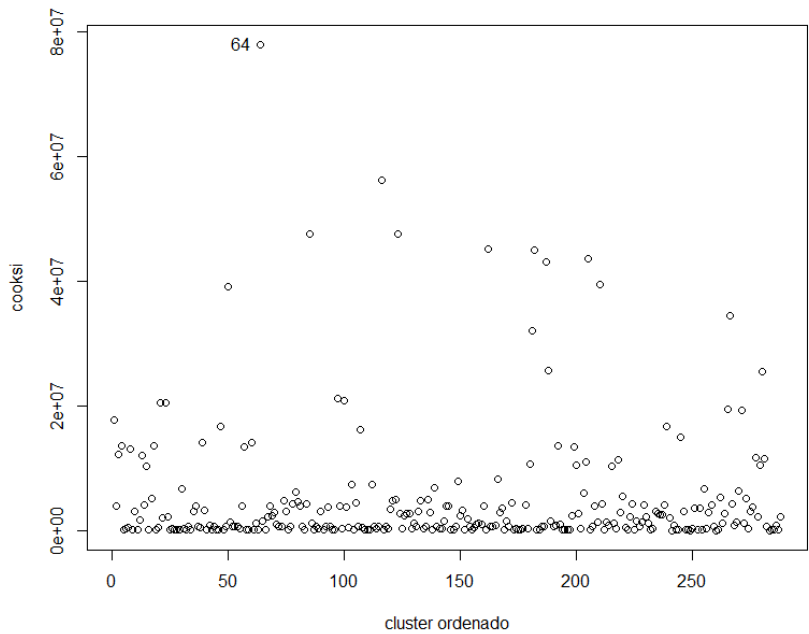
La segunda de las gráficas corresponde a la medida LeverageCi con matriz de correlación de trabajo independiente. Según esta medida no hay puntos leverages.

**GRÁFICA 11. MEDIDA LEVERAGE PARA SMOKE CON MATRIZ DE CORRELACIÓN DE TRABAJO INDEPENDIENTE**



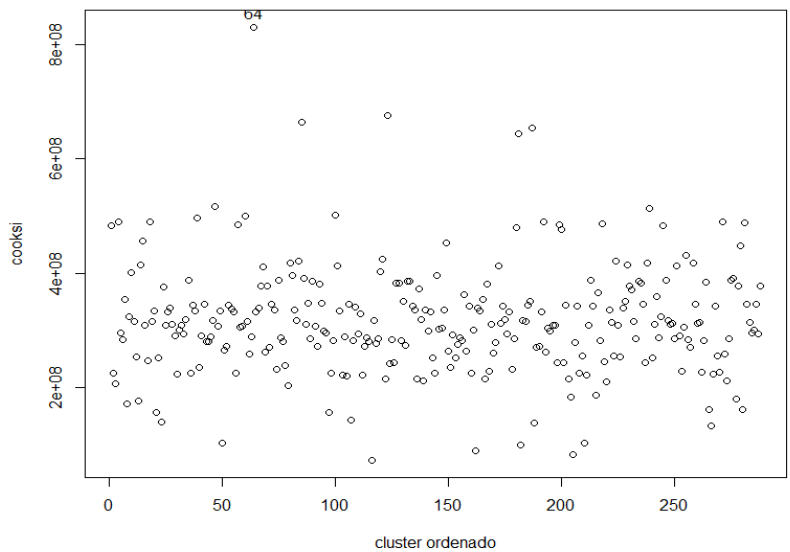
La tercera de las gráficas corresponde a la distancia de Cook con matriz de correlación de trabajo AR1. Según esta medida hay un individuo que influye más que el resto en los coeficientes de regresión: 64.

**GRÁFICA 12. DISTANCIA DE COOK PARA NEPAL CON MATRIZ DE CORRELACIÓN DE TRABAJO AR1**



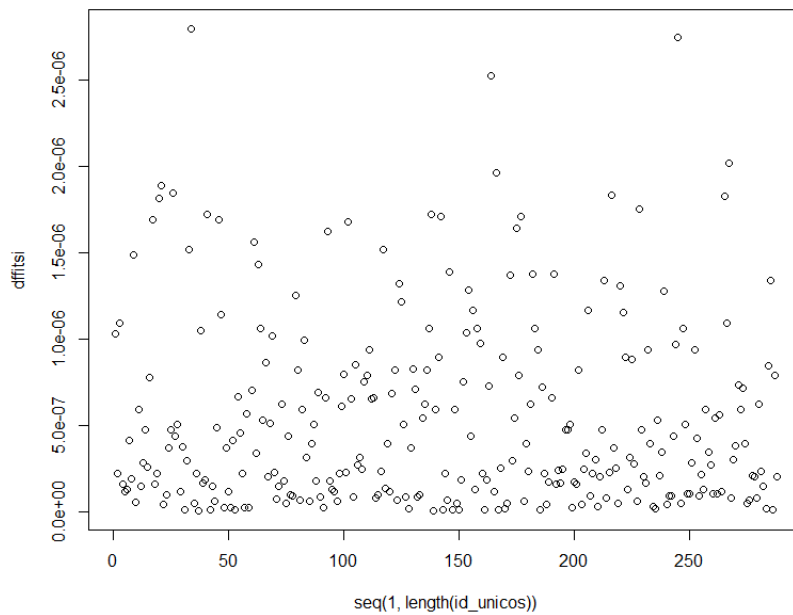
La cuarta de las gráficas corresponde a la distancia de Cook con matriz de correlación de trabajo independiente. Según esta medida hay un individuo que influye más que el resto en los coeficientes de regresión: 64.

**GRÁFICA 13. DISTANCIA DE COOK PARA NEPAL CON MATRIZ DE CORRELACIÓN DE TRABAJO INDEPENDIENTE**



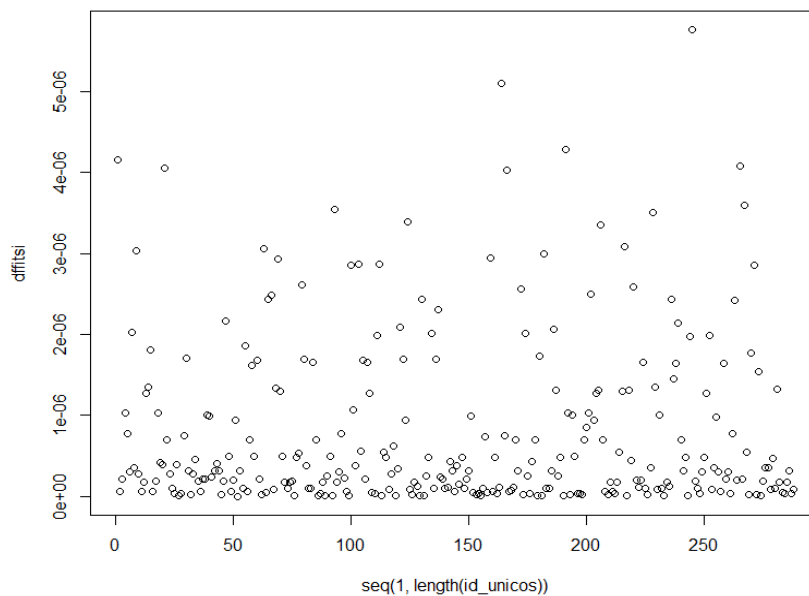
La quinta de las gráficas corresponde a la medida de influencia DFFITCi con matriz de correlación de trabajo AR1. Según esta medida no hay observaciones que influyan sobre el ajuste global.

**GRÁFICA 14. MEDIDA DE INFLUENCIA DFFITCi PARA SMOKE CON MATRIZ DE CORRELACIÓN DE TRABAJO AR1**



Por último, la sexta de las gráficas corresponde a la medida de influencia DFFITCi con matriz de correlación de trabajo independiente. Según esta medida no hay observaciones que influyan sobre el ajuste global.

**GRÁFICA 15. MEDIDA DE INFLUENCIA DFFITCi PARA SMOKE CON MATRIZ DE CORRELACIÓN DE TRABAJO INDEPENDIENTE**



Como podemos ver, las medidas de diagnóstico implementadas en las EEG detectan las mismas observaciones, incluso alguna más, que estas mismas medidas cuando usamos el MLG, como era de esperar.



