

# **Características epidemiológicas y clínicas de los Trastornos del Neurodesarrollo en población infantil en el periodo 2016 al 2023**

Gloria Andrea Rico Velasco  
Universidad San Buenaventura

Tuesday, December 3, 2024

## **Table of contents**

1 Instalacion de paquetes .....	1
2 Funciones para visualizacion de datos .....	1
3 Lectura y limpieza de dataset .....	2
3.1 Procesamiento de los datos .....	3
4 Exploración visual de los datos .....	3
5 Modelo de regresion lineal .....	5
5.1 Ajuste del modelo .....	6
6 Modelo multinivel .....	11

Esta investigación tiene como objetivo describir las características epidemiológicas de los trastornos del neurodesarrollo en población infantil en Colombia.

## **1 Instalacion de paquetes**

A continuación se instalan los paquetes necesarios para desarrollar el modelo estadístico.

```
if (!require("brms")) install.packages("brms")
if (!require("broom.mixed")) install.packages("broom.mixed")
if (!require("dplyr")) install.packages("dplyr")
if (!require("ggalluvial")) install.packages("ggalluvial")
if (!require("ggplot2")) install.packages("ggplot2")
if (!require("glmmTMB")) install.packages("glmmTMB")
if (!require("here")) install.packages("here")
if (!require("lme4")) install.packages("lme4")
if (!require("MASS")) install.packages("MASS")
if (!require("MetBrewer")) install.packages("MetBrewer")
if (!require("performance")) install.packages("performance")
if (!require("readxl")) install.packages("readxl")
if (!require("tidybayes")) install.packages("tidybayes")
```

## **2 Funciones para visualizacion de datos**

```

Plot_theme <- theme_classic() +
  theme(
    plot.title = element_text(size=18, hjust = 0.5, face="bold"),
    plot.subtitle = element_text(size = 10, color = "black"),
    plot.caption = element_text(size = 12, color = "black"),
    axis.line = element_line(colour = "black", linewidth = 1.5, linetype =
"solid"),
    axis.ticks.length=unit(7,"pt"),

    axis.title.x = element_text(colour = "black", size = 16),
    axis.text.x = element_text(colour = "black", size = 16, angle = 0, hjust =
0.5),
    axis.ticks.x = element_line(colour = "black", linewidth = 1),

    axis.title.y = element_text(colour = "black", size = 16),
    axis.text.y = element_text(colour = "black", size = 16),
    axis.ticks.y = element_line(colour = "black", linewidth = 1),

    legend.position="right",
    legend.direction="vertical",
    legend.title = element_text(colour="black", face="bold", size=12),
    legend.text = element_text(colour="black", size=10),

    plot.margin = margin(t = 10, # Top margin
                         r = 2, # Right margin
                         b = 10, # Bottom margin
                         l = 10) # Left margin
  )

```

### 3 Lectura y limpieza de dataset

Se cargan los datos contenidos en la tabla “Data/Ministerio\_Datos\_Limpios.xlsx”. Se eliminan las casillas con valores faltantes “NA”

```

##leer archivo excel

Datos_Limpios_Analisis_Ministerio           <-      read_excel(here("Data/
Ministerio_Datos_Limpios.xlsx"))

# Eliminar filas con valores NA en cualquier columna
Datos_Limpios_Analisis_Ministerio <- na.omit(Datos_Limpios_Analisis_Ministerio)

str (Datos_Limpios_Analisis_Ministerio)

```

```

tibble [772,313 × 8] (S3: tbl_df/tbl/data.frame)
$ Año          : num [1:772313] 2016 2016 2016 2016 2016 ...

```

```

$ Año_cod      : num [1:772313] 0 0 0 0 0 0 0 0 0 0 ...
$ Departamento: chr  [1:772313] "Antioquia" "Antioquia" "Antioquia"
"Antioquia" ...
$ Municipio   : chr [1:772313] "Medellín" "Medellín" "Medellín" "Medellín" ...
$ Diagnostico : chr [1:772313] "F700" "F700" "F700" "F700" ...
$ Edades       : chr [1:772313] "De 02 años" "De 04 años" "De 04 años" "De 06
años" ...
$ Sexo         : chr [1:772313] "MASCULINO" "FEMENINO" "MASCULINO" "FEMENINO" ...
$ cantidad_p   : num [1:772313] 1 2 1 1 1 3 8 5 6 8 ...
- attr(*, "na.action")= 'omit' Named int [1:14368] 13547 13548 13549 13550 13551
13552 13553 13554 13555 13556 ...
... attr(*, "names")= chr [1:14368] "13547" "13548" "13549" "13550" ...

```

### 3.1 Procesamiento de los datos

A continuación procesamos la columna de “Edades” para seleccionar solamente los valores numéricos y establecemos la columna de sexo como un factor:

```

Datos_Limpios_Analisis_Ministerio <- Datos_Limpios_Analisis_Ministerio %>%
  mutate(
    Edades = as.numeric(gsub("De | años", "", Edades)),
    Sexo = factor(Sexo, levels = c("FEMENINO", "MASCULINO")),
    Diagnostico = factor(Diagnostico))

str(Datos_Limpios_Analisis_Ministerio)

```

```

tibble [772,313 × 8] (S3: tbl_df/tbl/data.frame)
$ Año          : num [1:772313] 2016 2016 2016 2016 2016 ...
$ Año_cod      : num [1:772313] 0 0 0 0 0 0 0 0 0 0 ...
$ Departamento: chr  [1:772313] "Antioquia" "Antioquia" "Antioquia"
"Antioquia" ...
$ Municipio   : chr [1:772313] "Medellín" "Medellín" "Medellín" "Medellín" ...
$ Diagnostico : Factor w/ 50 levels "F700","F701",..., 1 1 1 1 1 1 1 1 1 1 ...
$ Edades       : num [1:772313] 2 4 4 6 6 7 7 8 8 9 ...
$ Sexo         : Factor w/ 2 levels "FEMENINO","MASCULINO": 2 1 2 1 2 1 2 1 2
1 ...
$ cantidad_p   : num [1:772313] 1 2 1 1 1 3 8 5 6 8 ...
- attr(*, "na.action")= 'omit' Named int [1:14368] 13547 13548 13549 13550 13551
13552 13553 13554 13555 13556 ...
... attr(*, "names")= chr [1:14368] "13547" "13548" "13549" "13550" ...

```

## 4 Exploración visual de los datos

Para este ejercicio analizaremos tres diagnósticos. Los datos utilizados, denominados Datos\_top\_3, corresponden a los diagnósticos más prevalentes de trastornos del neurodesarrollo. En primer

lugar, filtramos los datos, obteniendo los tres diagnósticos más comunes y asegurandonos de que la tabla contenga solamente filas con “FEMENINO” Y “MASCULINO.

```
top_3_diagnosticos <- Datos_Limpios_Analisis_Ministerio %>%
  group_by(Diagnstico) %>%
  summarise(total_diagnosticos = sum(cantidad_p, na.rm = TRUE)) %>%
  arrange(desc(total_diagnosticos)) %>%
  slice_head(n = 3) %>%
  pull(Diagnstico) # Extraer los 3 diagnósticos más prevalentes

Datos_top_3 <- Datos_Limpios_Analisis_Ministerio %>%
  filter(Diagnstico %in% top_3_diagnosticos)
Datos_top_3 <- filter(Datos_top_3, Sexo == 'FEMENINO' | Sexo == 'MASCULINO')
Datos_top_3$Diagnstico <- factor(Datos_top_3$Diagnstico, levels = c("F808",
  "F809", "F900"))
```

Ahora, visualizamos la distribución:

```
Neurodesarrollo_Expl <- ggplot(
  Datos_top_3 ,
  aes(
    x = Año,
    y = cantidad_p,
    color = Diagnstico
  )
) +
  geom_violin()+
  geom_point(position = position_jitter(width = 0.1)) +
  labs(y = "Cantidad de diagnósticos")+
  ggtitle("Evolución de los diagnósticos del neurodesarrollo")+
  Plot_theme +
  theme(legend.position = "bottom",
        legend.direction = "horizontal") +
  facet_wrap(~ Sexo)

ggsave(Neurodesarrollo_Expl,
       filename = here("Plots/Neurodesarrollo_Expl.png"),
       width     = 16,
       height    = 12,
       units     = "cm")
```

Neurodesarrollo\_Expl

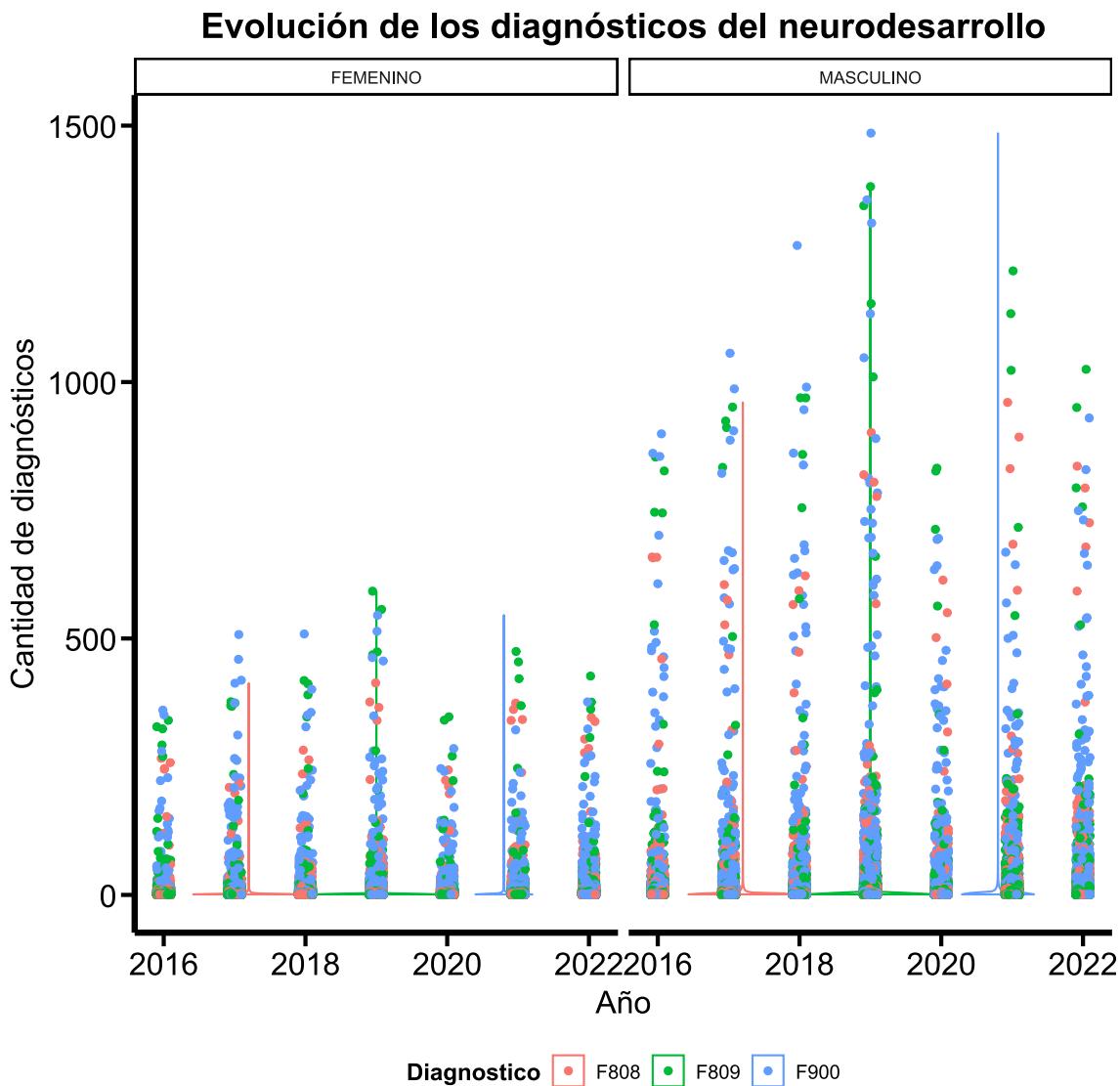


Figure 1: Evolución del número de diagnósticos por sexo

Figure 1 revela que los diagnósticos pueden alcanzar más de 1000 y presentar una dispersión sustancial que no permite observar su distribución. Esto indica que aunque existen cantidades elevadas de diagnósticos, la mayoría de ellos se encuentran centrados en números bajos (ceranos al cero). Debido a la naturaleza y dispersión de los datos, es oportuno modelarlos utilizando una distribución negativa binomial.

## 5 Modelo de regresión lineal

El objetivo de este análisis es evaluar el efecto de las variables Año\_cod (años codificados) y Diagnóstico (categoría de diagnóstico) sobre la variable dependiente cantidad\_p, que representa la cantidad de casos reportados. Ajustamos el siguiente modelo:

$$Casos = \beta_0 + \beta_1 \times Año + \beta_2 \times Diagnóstico + \beta_3 \times (Año \times Diagnóstico) + \epsilon_i$$

donde  $\beta_0$  es el intercepto, que representa el número de casos en el año (2016) y el diagnóstico base (F808).  $\beta_1$  es el efecto del año en el número de casos.  $\beta_2$  son los coefficients que representan el efecto de cada diagnóstico y  $\beta_3$  es la interacción entre los dos predictores. Finalmente,  $\epsilon_i$  representa el error que sigue una distribución student-t

## 5.1 Ajuste del modelo

Utilizamos una distribución binomial negativa (negbinomial) para modelar los datos, considerando la interacción entre el diagnóstico y el año.

```
Neurodesarrollo_Mdl1 <- bf(cantidad_p ~ Año_cod * Diagnostico)

get_prior(Neurodesarrollo_Mdl1, Datos_top_3, family = negbinomial())
```

prior	class	coef	group	resp	dpar	npar
(flat)	b					
(flat)	b		Año_cod			
(flat)	b	Año_cod:DiagnosticoF809				
(flat)	b	Año_cod:DiagnosticoF900				
(flat)	b		DiagnosticoF809			
(flat)	b		DiagnosticoF900			
student_t(3, 0, 2.5)	Intercept					
inv_gamma(0.4, 0.3)	shape					
lb ub	source					
	default					
	(vectorized)					
	(vectorized)					
	(vectorized)					
	(vectorized)					
	(vectorized)					
	default					
0	default					

```
Neurodesarrollo_Fit1 <-
  brm(
    data = Datos_top_3,
    family = negbinomial,
    formula = Neurodesarrollo_Mdl1,
    chains = 4,
    cores = 4,
    warmup = 2500,
    iter = 5000,
    seed = 8807,
    control = list(adapt_delta = 0.99, max_treedepth = 15),
    file = here("Models/Neurodesarrollo_Fit1.rds"),
```

```

    file_refit = "never"
)

```

A continuación, visualizamos los coefficients obtenidos:

```
summary(Neurodesarrollo_Fit1)
```

```

Family: negbinomial
Links: mu = log; shape = identity
Formula: cantidad_p ~ Año_cod * Diagnostico
Data: Datos_top_3 (Number of observations: 161220)
Draws: 4 chains, each with iter = 5000; warmup = 2500; thin = 1;
       total post-warmup draws = 10000

```

#### Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS
Intercept	1.35	0.01	1.33	1.37	1.00	3875
Año_cod	0.04	0.00	0.03	0.04	1.00	3725
DiagnosticoF809	0.12	0.02	0.09	0.15	1.00	3934
DiagnosticoF900	0.79	0.02	0.76	0.82	1.00	3783
Año_cod:DiagnosticoF809	-0.02	0.00	-0.03	-0.01	1.00	3757
Año_cod:DiagnosticoF900	-0.08	0.00	-0.09	-0.07	1.00	3634
			Tail_ESS			
Intercept		5296				
Año_cod		5515				
DiagnosticoF809		5497				
DiagnosticoF900		5000				
Año_cod:DiagnosticoF809		5350				
Año_cod:DiagnosticoF900		4702				

#### Further Distributional Parameters:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
shape	0.60	0.00	0.59	0.60	1.00	7695	6302

Draws were sampled using sampling(NUTS). For each parameter, Bulk\_ESS and Tail\_ESS are effective sample size measures, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat = 1).

Con el procedimiento anterior utilizamos una distribución binomial negativa para la variable respuesta. Esta distribución es apropiada en casos donde el recuento de unidades presenta sobre-dispersión, es decir, donde la varianza supera con creces a la media.

La función de enlace para el parámetro de la media ( $\mu$ ) es el enlace logarítmico. Esto implica que  $\log(\mu) = \text{predictor lineal}$ . Para interpretar estos resultados en la escala natural, es necesario exponentiar los coefficients:

- El intercepto (el valor del diagnóstico F808 en el 2016) es  $\exp(1.35) \approx 3.86$

- El efecto del año es  $\exp(0.04) \approx 1.04$ . Esto implica un aumento del 4% por cada año que pasa.
- El diagnóstico F809 hay un incremento de  $\exp(0.12) \approx 1.13$ , un aumento del 13% en el 2016. Por otra parte, el aumento para F900 aumenta en  $\exp(0.79) \approx 2.20$ , un incremento del 120%.
- Para el diagnóstico F809, el cambio por año reduce el incremento de diagnósticos en  $\exp(-0.02) \approx 0.98$ . Es decir, una disminución del 2% por unidad. Por el contrario, esta disminución es de  $\exp(-0.08) \approx 0.92$  (8% por unidad) para el diagnóstico F900.
- Finalmente, el coefficiente para shape es 0,60, lo que nos sugiere una sobredispersión moderada en los conteos.

Ahora visualizamos los resultados utilizando la función `conditional_effects`:

```
Neurodesarrollo_CE <- conditional_effects(Neurodesarrollo_Fit1)

Neurodesarrollo_CE_Fig <- plot(Neurodesarrollo_CE , plot = FALSE)[[3]]

Neurodesarrollo_CE_Fig <- Neurodesarrollo_CE_Fig +
  labs(title = "Evolución del número de diagnósticos",
       y = "Número de diagnósticos",
       x = "Año") +
  Plot_theme +
  theme(legend.position = "bottom",
        legend.direction = "horizontal")

ggsave (Neurodesarrollo_CE_Fig,
        file = here("Plots/Neurodesarrollo_Fit1.jpg"),
        width = 15,
        height = 12,
        units = "cm")

Neurodesarrollo_CE_Fig
```

## Evolución del número de diagnósticos

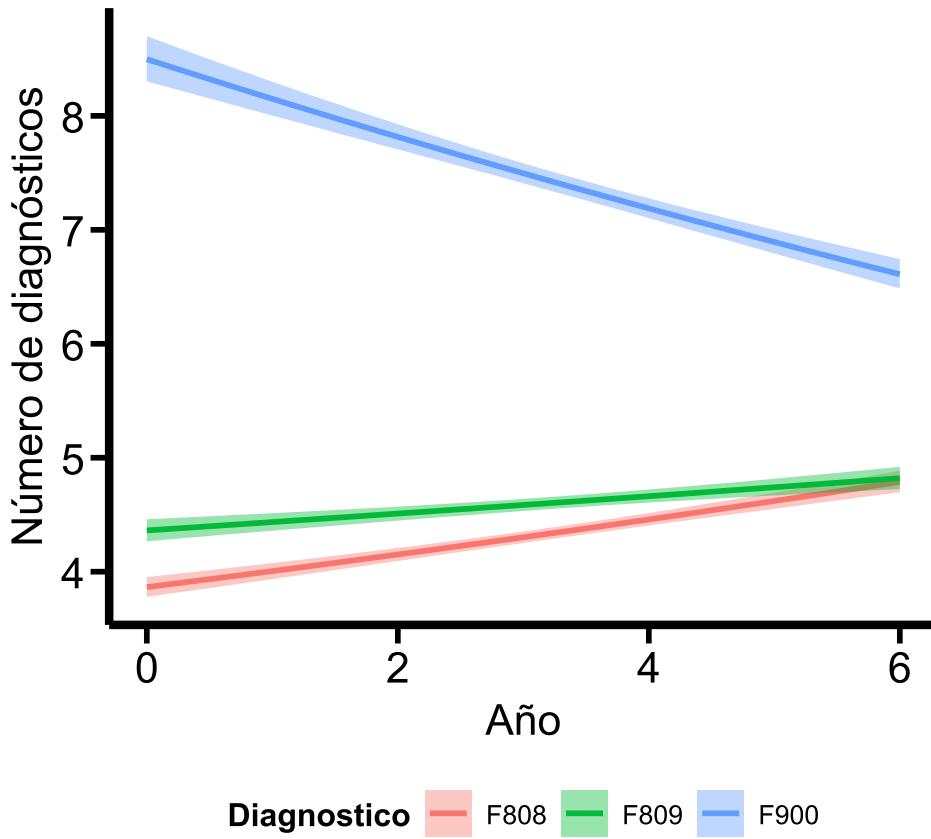


Figure 2: Evolución del número de diagnósticos por año

Figure 2 nos muestra las estimaciones como fueran descritas anteriormente. Podemos utilizar una visualización alternativo con la función `add_epred_draws` de `tidybayes`.

```
clrs <- MetBrewer::met.brewer("Lakota", 6)

Neurodesarrollo_CE_Fig2 <- Datos_top_3 %>%
  add_epred_draws(Neurodesarrollo_Fit1, ndraws = 50, allow_new_levels = TRUE) %>%
  ggplot(aes(x = Año_cod, y = cantidad_p, color = Diagnóstico)) +
  geom_point(data = Datos_top_3, size = 1) +
  geom_line(aes(y = .epred, group = paste(Diagnóstico, .draw)), size = 0.5, alpha = 0.3) +
  scale_color_manual(values = c(clrs[6], clrs[3], clrs[2])) +
  labs(title = "Evolución del número de diagnósticos",
       x = "Año",
       y = "Número de diagnósticos",
       color = "Party") +
```

```

scale_y_continuous(limits = c(0, 10)) +
Plot_theme +
theme(legend.position = "bottom",
      legend.direction = "horizontal")

ggsave (Neurodesarrollo_CE_Fig2,
        file = here("Plots/Neurodesarrollo_Fit1_2.jpg"),
        width = 15,
        height = 12,
        units = "cm")

```

Adicionalmente, podemos hacer una validación del modelo utilizando la función `pp_check` de `brms`.

```

pp_check(Neurodesarrollo_Fit1, ndraws = 500) +
  scale_x_continuous(limits=c(0, 25))

```

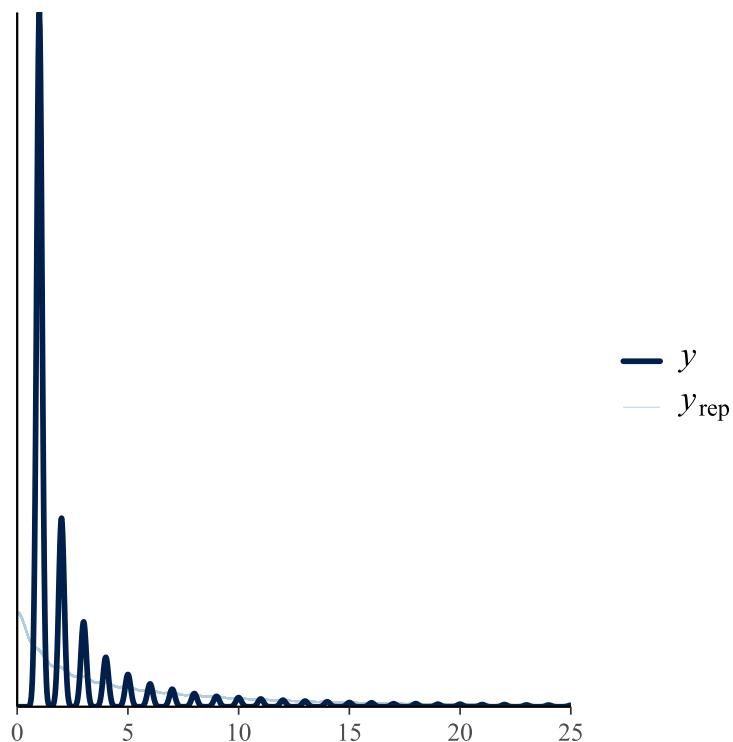


Figure 3: Evaluación del modelo ajustado

Figure 3 Nos muestra que nuestras predicciones se desvían sustancialmente de las observaciones. Esto quiere decir que nuestro modelo no es lo suficientemente explicativo para capturar la naturaleza del número de diagnósticos.

## 6 Modelo multinivel

Como complemento, realizamos el ajusted de un modelo multinivel para considerar la dispersión de datos por departamento.

```
Neurodesarrollo_Fit2 <- glmer(cantidad_p ~ Año_cod * Diagnostico + (1|  
Departamento),  
family = "poisson",  
data = Datos_top_3,  
control = glmerControl(optimizer = "bobyqa"))
```

Imprimimos los resultados de igual manera que con el modelo anterior:

```
summary(Neurodesarrollo_Fit2)
```

```
Generalized linear mixed model fit by maximum likelihood (Laplace  
Approximation) [glmerMod]  
Family: poisson ( log )  
Formula: cantidad_p ~ Año_cod * Diagnostico + (1 | Departamento)  
Data: Datos_top_3  
Control: glmerControl(optimizer = "bobyqa")  
  
AIC      BIC      logLik deviance df.resid  
2646418  2646488 -1323202   2646404    161213  
  
Scaled residuals:  
    Min     1Q Median     3Q    Max  
-10.772 -1.663 -1.079 -0.463 276.526  
  
Random effects:  
Groups      Name      Variance Std.Dev.  
Departamento (Intercept) 0.5112   0.715  
Number of obs: 161220, groups: Departamento, 33  
  
Fixed effects:  
            Estimate Std. Error z value Pr(>|z|)  
(Intercept) 0.907380  0.117929  7.694 1.42e-14 ***  
Año_cod      0.045770  0.001068 42.874 < 2e-16 ***  
DiagnosticoF809 0.068892  0.005632 12.232 < 2e-16 ***  
DiagnosticoF900  0.822469  0.005236 157.068 < 2e-16 ***  
Año_cod:DiagnosticoF809 -0.003770  0.001454 -2.592  0.00953 **  
Año_cod:DiagnosticoF900 -0.104984  0.001362 -77.097 < 2e-16 ***  
---  
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### Correlation of Fixed Effects:

```
(Intr) Año_cd DgF809 DgF900 Añ_:DF8
Año_cod      -0.035
DignstcF809 -0.032  0.639
DignstcF900 -0.033  0.687  0.584
Añ_cd:DF809  0.028 -0.732 -0.859 -0.503
Añ_cd:DF900  0.029 -0.782 -0.500 -0.863  0.573
optimizer (bobyqa) convergence code: 0 (OK)
Model failed to converge with max|grad| = 0.00518234 (tol = 0.002, component 1)
Model is nearly unidentifiable: very large eigenvalue
- Rescale variables?
```

Y visualizamos los resultados:

```
tidy_re <- tidy(Neurodesarrollo_Fit2 , effects = "ran_vals", conf.int = TRUE)

tidy_Departamento <- tidy_re %>%
  filter(group == "Departamento")

ggplot(tidy_Departamento, aes(x = estimate, y = level)) +
  geom_point() +
  geom_errorbarh(aes(xmin = conf.low, xmax = conf.high)) +
  labs(
    title = "Group-level effects por departamento",
    x = "Caso",
    y = "Departamento"
  ) +
  theme_minimal()
```

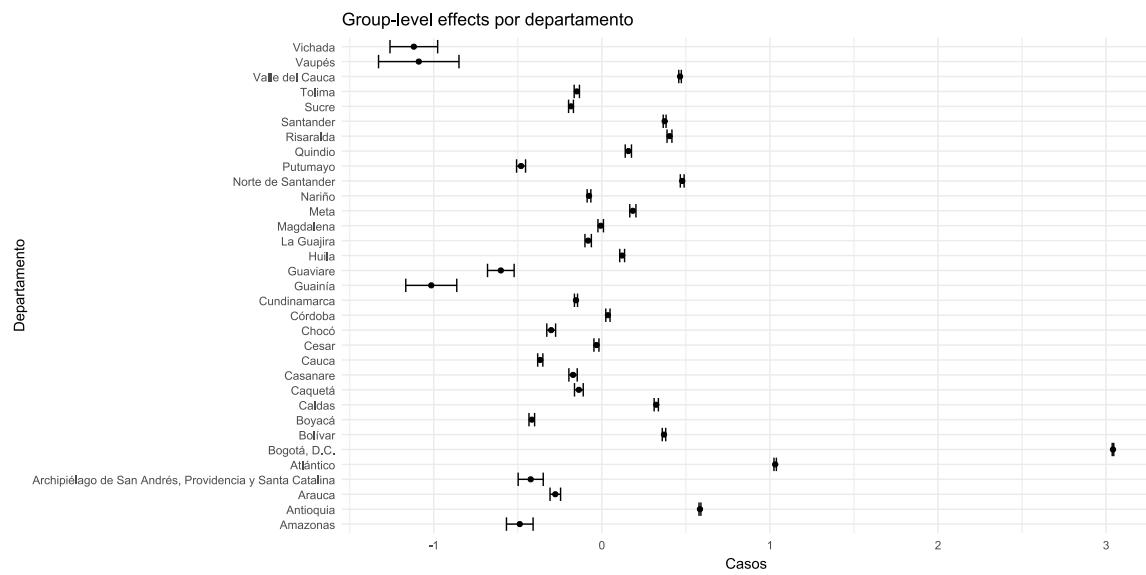


Figure 4: Evaluación del modelo ajustado

Figure 4 nos muestra la desviación de diagnósticos por departamento.