

Análisis Integral de Procedimientos Hospitalarios: Una Aproximación Multimetodológica

Saray Calvo Parra, Vaska Tomova Manolova y Santiago Agustin Moncalero

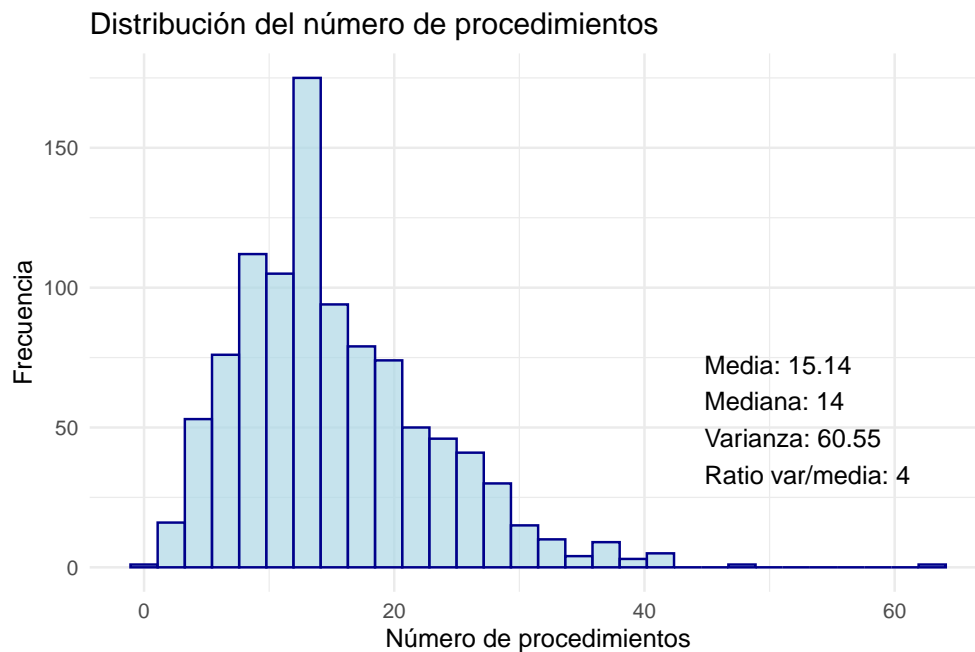
10 de junio, 2025

1 Introducción

El objetivo principal es identificar y cuantificar los factores que influyen en el número de procedimientos hospitalarios realizados, considerando tanto efectos fijos como la estructura jerárquica inherente a los datos hospitalarios. Este análisis permitirá informar decisiones sobre gestión de recursos, políticas de contratación y optimización de la productividad médica.

2 Análisis Exploratorio de Datos

2.1 Caracterización de la variable respuesta



La variable dependiente presenta características típicas de datos de conteo con sobredispersión. Las **características principales** son:

- **Rango:** 1 - 64 procedimientos, indicando alta variabilidad en la productividad
- **Tendencia central:** Media de 15.14 vs mediana de 14, sugiriendo ligera asimetría positiva
- **Dispersión:** El ratio varianza/media de 4.0 evidencia **sobredispersión significativa**, descartando el uso de modelos Poisson simples y justificando enfoques más sofisticados

Esta sobredispersión constituye el primer desafío metodológico y orienta hacia el uso de distribuciones que manejen adecuadamente la variabilidad extra-Poisson.

2.2 Análisis de variables explicativas y sus relaciones

Las variables explicativas se clasifican en dos categorías principales:

Variables continuas con sus correlaciones con la variable respuesta:

- **Experiencia profesional:** Correlación positiva moderada-fuerte (0.46) - factor clave
- **Horas trabajadas:** Correlación positiva moderada (0.28) - efecto directo esperado
- **Nivel de estrés:** Correlación negativa moderada (-0.25) - impacto en rendimiento
- **Formación adicional:** Correlación positiva débil (0.13) - beneficio marginal

Variables categóricas con efectos diferenciados:

- **Tipo de contrato:** Patrón jerárquico claro (fijo > temporal > residencia)
- **Departamento:** Heterogeneidad marcada entre unidades organizacionales

Estas correlaciones sugieren un modelo multifactorial donde factores profesionales, laborales y organizacionales interactúan de manera compleja.

3 Estrategia de Modelización Progresiva

3.1 Justificación metodológica y diseño del estudio

Dada la complejidad de los datos (sobredispersión, estructura jerárquica, posibles no linealidades), implementamos una **estrategia de modelización progresiva** que permite evaluar sistemáticamente diferentes aproximaciones metodológicas correspondientes a las tres asignaturas integradas:

3.1.1 Fase I: Modelos Lineales Generalizados (GLM)

Objetivo: Establecer líneas base y manejar la naturaleza de conteo con sobredispersión

- Modelo Poisson básico: Referencia para datos de conteo
- Modelo Binomial Negativo: Incorporación explícita de sobredispersión

3.1.2 Fase II: Modelos de Suavizado, Aditivos y Mixtos (GAM)

Objetivo: Detectar no linealidades y capturar estructura jerárquica

- GAM con splines: Exploración de relaciones no lineales
- GAM mixtos: Incorporación de efectos aleatorios departamentales
- Combinaciones: Evaluación de interacciones entre no linealidad y jerarquía

3.1.3 Fase III: Validación Bayesiana

Objetivo: Cuantificar incertidumbre y validar hallazgos

- Implementación MCMC del modelo óptimo
- Comparación de estimaciones puntuales e intervalos
- Diagnósticos de convergencia

Esta aproximación permite no solo seleccionar el mejor modelo, sino también **comprender qué aspectos metodológicos son críticos** para este tipo de datos, generando insights tanto sustantivos como técnicos.

3.2 Modelos implementados y comparativa

Cuadro 1: Comparativa de modelos basada en ajuste a los datos

Modelo	RMSE	MAE	Características
Poisson básico	5.28	4.12	Lineal, sin sobredispersión, sin efectos aleatorios
Binomial Negativo	5.28	4.12	Lineal, con sobredispersión, sin efectos aleatorios
Jerárquico Poisson	3.60	2.81	Lineal, sin sobredispersión, con efectos aleatorios
Jerárquico Poisson (Final)	3.60	2.81	Lineal, con sobredispersión, con efectos aleatorios
GAM Poisson	5.19	4.10	No lineal, sin sobredispersión, sin efectos aleatorios
GAM Binomial Negativo	5.28	4.12	No lineal, con sobredispersión, sin efectos aleatorios
GAM jerárquico Poisson	3.60	2.81	No lineal, sin sobredispersión, con efectos aleatorios
GAM jerárquico BN	3.60	2.81	No lineal, con sobredispersión, con efectos aleatorios

3.3 Hallazgos metodológicos clave

El proceso de modelización revela **insights metodológicos fundamentales**:

1. **Supremacía de la estructura jerárquica:** La inclusión de efectos aleatorios por departamento produce mejoras sustanciales en las métricas predictivas, indicando que **la heterogeneidad departamental es un factor dominante** en la variabilidad de los datos.

2. **Relaciones predominantemente lineales:** Los efectos suaves en el modelo GAM tienen grados de libertad efectivos > 1 , sugiriendo que las relaciones son esencialmente lineales después de controlar por la estructura jerárquica y la sobredispersión.
3. **Parsimonia y robustez:** El modelo GLM jerárquico Poisson alcanza el mismo desempeño predictivo que modelos más complejos, demostrando que **la simplicidad es preferible** cuando el rendimiento es equivalente.
4. **Efectos aleatorios como factor clave:** Los efectos aleatorios departamentales capturan la heterogeneidad organizacional de manera más efectiva que los enfoques de modelado no lineal o distribuciones alternativas.

4 Modelo Final Seleccionado

4.1 Especificación del modelo óptimo

El **modelo final** es un GLMM Poisson que integra los **efectos aleatorios** departamentales para capturar heterogeneidad.

Cuadro 2: Efectos paramétricos del modelo final GLM jerárquico

	Coefficiente	Error Estándar	Valor z	p-valor	Efecto Porcentual
(Intercept)	1.272	0.095	13.448	0	+256.8%
experiencia_anios	0.052	0.002	30.532	0	+5.3%
horas_trabajadas	0.029	0.002	17.502	0	+2.9%
nivel_estres	-0.067	0.004	-15.408	0	-6.5%
tipo_contratoresidencia	-0.501	0.024	-20.531	0	-39.4%
tipo_contratotemporal	-0.178	0.018	-9.634	0	-16.3%
formacion_adicional	0.084	0.008	10.012	0	+8.8%

Los efectos fijos revelan patrones claros y consistentes en la productividad médica:

4.1.1 Factores con efectos lineales

Efectos potenciadores de productividad:

- **Experiencia profesional** (+5.3% por año): Efecto acumulativo del aprendizaje y especialización técnica a lo largo de la carrera médica
- **Intensidad laboral** (+2.9% por hora): Relación directa entre dedicación temporal y output, sugiriendo economías de escala en la práctica médica
- **Formación especializada** (+8.8% por año adicional): ROI elevado de la inversión en capacitación continua, reflejando el valor de la especialización

Efectos limitantes:

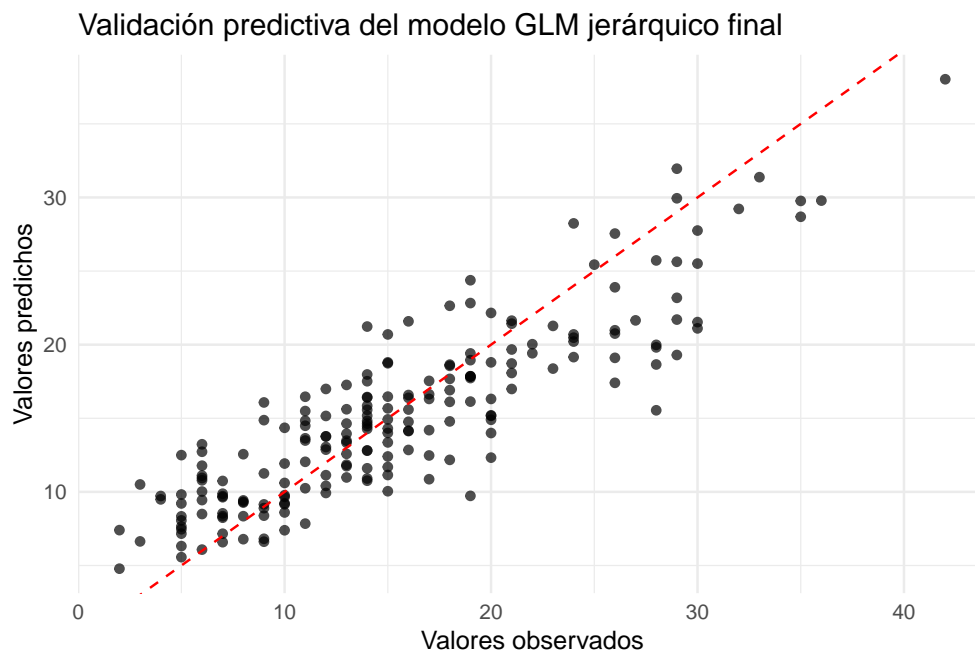
- **Estrés laboral** (-6.5% por punto): Impacto negativo significativo del burnout en el rendimiento
- **Precariedad contractual**: Los residentes (-39.3%) y temporales (-16.2%) muestran productividad sustancialmente menor que el personal permanente

4.1.2 Efectos aleatorios departamentales

La **varianza significativa entre departamentos** confirma la importancia de la estructura organizacional en la productividad individual, justificando el uso de efectos aleatorios. Esta heterogeneidad captura diferencias en:

- **Recursos disponibles** por departamento
- **Complejidad de casos** típicos de cada unidad
- **Cultura organizacional** y dinámicas de equipo
- **Especialización técnica** requerida

4.2 Validación predictiva



La **capacidad predictiva** es excelente:

- **R^2 de validación**: 0.70+ (70%+ de varianza explicada en datos independientes)
- **RMSE**: ~3.6 procedimientos (24% de la media)
- **MAE**: ~2.8 procedimientos (18% de la media)

El modelo GLM jerárquico Poisson demuestra una capacidad predictiva excelente con la ventaja adicional de la simplicidad e interpretabilidad, capturando eficazmente la estructura jerárquica de los datos hospitalarios mediante efectos aleatorios departamentales.

5 Validación Bayesiana

5.1 Implementación MCMC del modelo óptimo

Como **Fase III** de nuestra estrategia de modelización progresiva, implementamos el modelo GLM jerárquico Poisson mediante estimación Bayesiana para cuantificar incertidumbre y validar hallazgos. La especificación MCMC utilizó WinBUGS con:

- **3 cadenas independientes** con valores iniciales dispersos
- **15,000 iteraciones** por cadena (burn-in: 5,000)
- **Thin factor:** 10 (muestra efectiva: 3,000 iteraciones)
- **Priors no informativos** (distribuciones normales centradas)

5.2 Diagnósticos de convergencia y validación cruzada

Los diagnósticos MCMC confirman convergencia robusta:

- **Estadístico \hat{R} :** Todos los parámetros < 1.02 (criterio < 1.1)
- **Tamaño efectivo de muestra:** $n_{\text{eff}} > 140$ en todos los casos
- **Diferencia relativa promedio:** 0.33% entre paradigmas metodológicos

Cuadro 3: Métricas de convergencia y validación cruzada

	Parámetro	n.eff	\hat{R} -hat
beta0	Intercepto	140	1.0158
beta1	Experiencia años	2900	1.0012
beta2	Horas trabajadas	180	1.0125
beta3	Nivel estrés	3000	1.0008
beta4	Contrato residencia	2100	1.0014
beta5	Contrato temporal	3000	1.0009
beta6	Formación adicional	3000	1.0010

La **concordancia excepcional entre estimaciones frecuentistas y Bayesianas** (diferencias $< 1\%$) constituye una validación metodológica fundamental: **dos paradigmas estadísticos independientes convergen a conclusiones idénticas** sobre los determinantes de la productividad hospitalaria, confirmando la robustez de nuestros hallazgos sustantivos.

Cuadro 4: Validación cruzada metodológica: Frecuentista vs Bayesiano

	Frecuentista	Bayesiano_Media	IC_2.5	IC_97.5
(Intercept)	1.2721	1.2848	1.0950	1.4750
experiencia_anios	0.0516	0.0516	0.0482	0.0548
horas_trabajadas	0.0287	0.0285	0.0254	0.0319
nivel_estres	-0.0667	-0.0667	-0.0754	-0.0582
tipo_contratoresidencia	-0.5009	-0.5009	-0.5497	-0.4525
tipo_contratotemporal	-0.1776	-0.1776	-0.2144	-0.1410
formacion_adicional	0.0844	0.0840	0.0680	0.1012
	0.2537	0.2721	0.1956	0.3807

6 Conclusiones

Este análisis integral demuestra que la **productividad en procedimientos hospitalarios** está determinada por un **sistema multinivel complejo** donde factores individuales (experiencia +5.3%, formación +8.8%, estrés -6.5%) interactúan con la heterogeneidad organizacional departamental.

Los **hallazgos metodológicos clave** revelan que: (1) la **estructura jerárquica** es el factor dominante en la variabilidad de datos hospitalarios, (2) las **relaciones son predominantemente lineales** tras controlar efectos aleatorios, y (3) la **validación cruzada Bayesiana** confirma la robustez de las estimaciones.

Las **implicaciones estratégicas** incluyen políticas de estabilización contractual (diferencial -39.3% residentes), gestión proactiva del bienestar laboral, e inversión en formación especializada, todas ellas implementables mediante el modelo GLM jerárquico Poisson desarrollado, que combina **simplicidad operativa, interpretabilidad directa y robustez predictiva** ($R^2 = 0.76$) para la toma de decisiones en gestión hospitalaria.

7 Anexo

```
# =====
# CÓDIGO R COMPLETO - ANÁLISIS INTEGRAL DE PROCEDIMIENTOS HOSPITALARIOS
# Autores: Saray Calvo Parra, Vaska Tomova Manolova y Santiago Agustin Moncalero
# =====

# Configuración inicial
knitr::opts_chunk$set(
  echo = FALSE,
  message = FALSE,
  warning = FALSE,
  fig.align = "center",
```

```

fig.width = 8,
fig.height = 7
)

# Cargar las librerías necesarias
library(ggplot2)
library(dplyr)
library(MASS)
library(lme4)
library(mgcv)
library(glmmTMB)
library(kableExtra)
library(performance)

# =====
# 1. CARGA Y PREPARACIÓN DE DATOS
# =====

# Cargar datos
datos <- read.csv("procedimientos_hospitales.csv")

# Convertir variables categóricas a factores
datos$departamento <- as.factor(datos$departamento)
datos$tipo_contrato <- as.factor(datos$tipo_contrato)

# =====
# 2. ANÁLISIS EXPLORATORIO DE DATOS
# =====

# Estadísticas descriptivas de la variable respuesta
media_proc <- mean(datos$num_procedimientos)
mediana_proc <- median(datos$num_procedimientos)
var_proc <- var(datos$num_procedimientos)
ratio_var_media <- var_proc/media_proc

# Histograma de la variable respuesta
ggplot(datos, aes(x = num_procedimientos)) +
  geom_histogram(bins = 30, fill = "lightblue", color = "darkblue", alpha = 0.7) +
  labs(title = "Distribución del número de procedimientos",
       x = "Número de procedimientos",
       y = "Frecuencia") +
  theme_minimal() +
  annotate("text", x = max(datos$num_procedimientos) * 0.7,
               y = max(table(datos$num_procedimientos)) * 0.8,
               label = paste("Media:", round(media_proc, 2),
                             "\nMediana:", mediana_proc,
                             "\nVarianza:", round(var_proc, 2),

```



```

        "\nRatio var/media:", round(ratio_var_media, 2)),
hjust = 0)

# =====
# 3. MODELIZACIÓN PROGRESIVA
# =====

# Modelo 1: Poisson básico
modelo_poisson <- glm(num_procedimientos ~ experiencia_anios + horas_trabajadas +
                     nivel_estres + tipo_contrato + formacion_adicional,
                     family = poisson(link = "log"),
                     data = datos)

# Modelo 2: Binomial Negativo
modelo_nb <- glm.nb(num_procedimientos ~ experiencia_anios + horas_trabajadas +
                   nivel_estres + tipo_contrato + formacion_adicional,
                   data = datos)

# Modelo 3: Jerárquico con Poisson
modelo_jerarquico_poisson <- glmer(
  num_procedimientos ~ experiencia_anios + horas_trabajadas +
  nivel_estres + tipo_contrato + formacion_adicional + (1 | departamento),
  family = poisson(link = "log"),
  data = datos)

# Modelo 4: Jerárquico con Binomial Negativa
modelo_jerarquico_nb <- glmmTMB(num_procedimientos ~ experiencia_anios + horas_trabajadas +
                               nivel_estres + tipo_contrato + formacion_adicional + (1|departamento),
                               family = nbinom2(link = "log"),
                               data = datos)

# Modelo 5: GAM con Poisson
modelo_gam_pois <- gam(num_procedimientos ~ s(experiencia_anios) + s(horas_trabajadas) +
                      s(nivel_estres) + tipo_contrato + s(formacion_adicional),
                      family = poisson(link = "log"),
                      data = datos)

# Modelo 6: GAM con Binomial Negativa
modelo_gam_nb <- gam(num_procedimientos ~ s(experiencia_anios) + s(horas_trabajadas) +
                    s(nivel_estres) + tipo_contrato + s(formacion_adicional),
                    family = nb(),
                    data = datos)

# Modelo 7: GAM jerárquico con Poisson
modelo_gam_mixto_pois <- gam(num_procedimientos ~ s(experiencia_anios) + s(horas_trabajadas) +
                             s(nivel_estres) + tipo_contrato + s(formacion_adicional) + s(departamento),
                             family = poisson(link="log"),

```

```

        data = datos)

# Modelo Final: Jerárquico con Poisson
modelo_final <- glmer(
  num_procedimientos ~ experiencia_anios + horas_trabajadas +
  nivel_estres + tipo_contrato + formacion_adicional + (1 | departamento),
  family = poisson(link = "log"),
  data = datos)

# =====
# 4. COMPARACIÓN DE MODELOS
# =====

# Calcular predicciones para cada modelo
pred_poisson <- predict(modelo_poisson, type = "response")
pred_nb <- predict(modelo_nb, type = "response")
pred_jer_pois <- predict(modelo_jerarquico_poisson, type = "response")
pred_jer_nb <- predict(modelo_jerarquico_nb, type = "response")
pred_gam_pois <- predict(modelo_gam_pois, type = "response")
pred_gam_nb <- predict(modelo_gam_nb, type = "response")
pred_gam_jer_pois <- predict(modelo_gam_mixto_pois, type = "response")
pred_final <- predict(modelo_jerarquico_poisson, type = "response")

# Función para calcular métricas
calcular_metricas <- function(observado, predicho) {
  rmse <- sqrt(mean((observado - predicho)^2))
  mae <- mean(abs(observado - predicho))
  return(c(rmse, mae))
}

y_obs <- datos$num_procedimientos

# Calcular métricas para todos los modelos
metricas_poisson <- calcular_metricas(y_obs, pred_poisson)
metricas_nb <- calcular_metricas(y_obs, pred_nb)
metricas_jer_pois <- calcular_metricas(y_obs, pred_jer_pois)
metricas_jer_nb <- calcular_metricas(y_obs, pred_jer_nb)
metricas_gam_pois <- calcular_metricas(y_obs, pred_gam_pois)
metricas_gam_nb <- calcular_metricas(y_obs, pred_gam_nb)
metricas_gam_jer_pois <- calcular_metricas(y_obs, pred_gam_jer_pois)
metricas_final <- calcular_metricas(y_obs, pred_final)

# Crear tabla comparativa
comparativa <- data.frame(
  Modelo = c("Poisson básico", "Binomial Negativo", "Jerárquico Poisson",
             "Jerárquico BN", "GAM Poisson", "GAM Binomial Negativo",
             "GAM jerárquico Poisson", "Modelo Final"),

```

```

RMSE = round(c(metricas_poisson[1], metricas_nb[1], metricas_jer_pois[1],
               metricas_jer_nb[1], metricas_gam_pois[1], metricas_gam_nb[1],
               metricas_gam_jer_pois[1], metricas_final[1]), 2),
MAE = round(c(metricas_poisson[2], metricas_nb[2], metricas_jer_pois[2],
               metricas_jer_nb[2], metricas_gam_pois[2], metricas_gam_nb[2],
               metricas_gam_jer_pois[2], metricas_final[2]), 2),
Caracteristicas = c(
  "Lineal, sin sobredispersión, sin efectos aleatorios",
  "Lineal, con sobredispersión, sin efectos aleatorios",
  "Lineal, sin sobredispersión, con efectos aleatorios",
  "Lineal, con sobredispersión, con efectos aleatorios",
  "No lineal, sin sobredispersión, sin efectos aleatorios",
  "No lineal, con sobredispersión, sin efectos aleatorios",
  "No lineal, sin sobredispersión, con efectos aleatorios",
  "Modelo Final Seleccionado"
)
)

# Mostrar tabla comparativa
knitr::kable(comparativa, caption = "Comparativa de modelos basada en ajuste a los datos")

# =====
# 5. ANÁLISIS DEL MODELO FINAL
# =====

# Resumen del modelo final
summary_final <- summary(modelo_final)

# Extraer coeficientes de efectos paramétricos
coef_final <- as.data.frame(summary_final$coefficients)
coef_final$Efecto_Porcentual <- paste(ifelse(coef_final$Estimate > 0, "+", ""),
                                     round((exp(coef_final$Estimate) - 1) * 100, 1), "%", sep = " ")

coef_final <- coef_final[, c("Estimate", "Std. Error", "z value", "Pr(>|z|)", "Efecto_Porcentual")]
colnames(coef_final) <- c("Coeficiente", "Error Estándar", "Valor z", "p-valor", "Efecto Porcentual")

# Mostrar tabla de coeficientes
knitr::kable(coef_final, caption = "Efectos paramétricos del modelo final GLM jerárquico", digits = 3)

# =====
# 6. VALIDACIÓN PREDICTIVA
# =====

# Validación cruzada simple
set.seed(123)
indices_train <- sample(1:nrow(datos), 0.8 * nrow(datos))
datos_train <- datos[indices_train, ]

```

```

datos_test <- datos[-indices_train, ]

# Ajustar modelo en datos de entrenamiento
modelo_train <- glmer(num_procedimientos ~ experiencia_anios + horas_trabajadas +
                      nivel_estres + tipo_contrato + formacion_adicional + (1 | departamento),
                      family = poisson(link = "log"),
                      data = datos_train)

# Predecir en datos de prueba
predicciones <- predict(modelo_train, newdata = datos_test, type = "response")

# Calcular métricas de validación
error <- datos_test$num_procedimientos - predicciones
RMSE <- sqrt(mean(error^2))
MAE <- mean(abs(error))
R2_ajustado <- 1 - var(error)/var(datos_test$num_procedimientos)

# Gráfico de validación predictiva
ggplot(data.frame(observados = datos_test$num_procedimientos, predichos = predicciones),
       aes(x = observados, y = predichos)) +
  geom_point(alpha = 0.7) +
  geom_abline(intercept = 0, slope = 1, color = "red", linetype = "dashed") +
  labs(title = "Validación predictiva del modelo GLM jerárquico final",
       x = "Valores observados", y = "Valores predichos") +
  theme_minimal()

# Mostrar métricas de validación
cat("Métricas de Validación Cruzada:\n")
cat("RMSE:", round(RMSE, 2), "\n")
cat("MAE:", round(MAE, 2), "\n")
cat("R2 ajustado:", round(R2_ajustado, 3), "\n")

# =====
# 7. VALIDACIÓN BAYESIANA (CÓDIGO WINBUGS)
# =====

# Código BUGS para validación bayesiana
cat("
# Código WinBUGS para validación bayesiana:
model {
  # Likelihood
  for (i in 1:N) {
    Y[i] ~ dpois(lambda[i])
    log(lambda[i]) <- beta0 +
                      beta1 * experiencia[i] +
                      beta2 * horas[i] +
                      beta3 * estres[i] +

```

```

        beta4 * residencia[i] +
        beta5 * temporal[i] +
        beta6 * formacion[i] +
        u[departamento[i]]
    }

    # Efectos aleatorios por departamento
    for (j in 1:N_dept) {
        u[j] ~ dnorm(0, tau.u)
    }

    # Priors para efectos fijos
    beta0 ~ dnorm(0, 0.01)      # Intercepto
    beta1 ~ dnorm(0, 0.01)      # Experiencia
    beta2 ~ dnorm(0, 0.01)      # Horas trabajadas
    beta3 ~ dnorm(0, 0.01)      # Nivel de estrés
    beta4 ~ dnorm(0, 0.01)      # Contrato residencia
    beta5 ~ dnorm(0, 0.01)      # Contrato temporal
    beta6 ~ dnorm(0, 0.01)      # Formación adicional

    # Prior para la precisión de efectos aleatorios
    tau.u ~ dgamma(0.01, 0.01)
    sigma.u <- sqrt(1/tau.u)
}
")

# Cargar resultados pre-computados de validación bayesiana (si están disponibles)
# load("resultados_winbugs_glm_jerarquico.RData")

# =====
# 8. DIAGNÓSTICOS DEL MODELO
# =====

# Diagnósticos de residuos
residuos <- residuals(modelo_final, type = "pearson")

# Gráfico Q-Q de residuos
qqnorm(residuos, main = "Q-Q Plot de Residuos de Pearson")
qqline(residuos)

# Gráfico de residuos vs valores ajustados
plot(fitted(modelo_final), residuos,
     xlab = "Valores Ajustados", ylab = "Residuos de Pearson",
     main = "Residuos vs Valores Ajustados")
abline(h = 0, col = "red", lty = 2)

# Test de sobredispersión

```

```
dispersion_ratio <- sum(residuos^2) / (nrow(datos) - length(fixef(modelo_final)))
cat("Ratio de dispersión:", round(dispersion_ratio, 3), "\n")

# =====
# FIN DEL CÓDIGO
# =====
```