

# TP 2: Modelos mixtos, splines penalizados y causalidad

Nicolás Celie, Martín Peralta, Nicolás Ian Rozenberg

2025-07-01

## Ejercicio 1: EDA

```
df_credits_train <- read.csv("credits_train.csv")
df_titles_train <- read.csv("titles_train.csv")

df_titles_train <- df_titles_train %>% # TODO: Chequear porcentaje de eliminados
  mutate(
    country = str_extract(production_countries, "[A-Z]{2}")
  ) %>%
  filter(!is.na(imdb_score) & !is.na(country))

df_titles_train$country <- as.factor(df_titles_train$country)

#head(df_credits_train)

#head(df_titles_train)
```

## Ejercicio 2

(a) Efectos fijos

```
fixed_model <- lm(imdb_score ~ country - 1, data = df_titles_train)
```

(b) Efectos aleatorios

```
random_model <- lmer(imdb_score ~ 1 + (1 | country), data = df_titles_train)
```

(c) Comparación

```
# Hacemos un conteo de títulos por país para visualizarlo en el gráfico
country_counts <- df_titles_train %>%
  count(country, name = "n")
```

```

fixed_estimates <- coef(fixed_model)
fixed_df <- tibble(
  country = names(fixed_estimates) %>% str_remove("^country"),
  fixed_effect = as.numeric(fixed_estimates)
)

random_ranef <- ranef(random_model)$country[, 1]
random_df <- tibble(
  country = rownames(ranef(random_model)$country),
  random_effect = as.numeric(random_ranef + fixef(random_model)[1])
)

comparison_df <- left_join(fixed_df, random_df, by = "country") %>%
  left_join(country_counts, by = "country") # Agregamos el conteo

comparison_df_long <- comparison_df %>%
  pivot_longer(cols = c("fixed_effect", "random_effect"),
    names_to = "model_type", values_to = "estimate")

ggplot(comparison_df_long, aes(x = reorder(country, estimate), y = estimate,
  color = model_type)) +
  geom_point(aes(size = n), alpha = 0.8) +
  geom_line(aes(group = country), color = "gray60", linetype = "dashed", size = 0.5) +
  labs(title = "Efectos Fijos vs. Aleatorios por Pais",
    x = "Pais", y = "Estimacion IMDB",
    color = "Tipo de modelo", size = "Cantidad de titulos") +
  scale_color_manual(values = c("fixed_effect" = "blue", "random_effect" = "red"),
    labels = c("Efectos Fijos", "Efectos Aleatorios")) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

```

```

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

```

## Efectos Fijos vs. Aleatorios por Pais

