INSTITUTO DE SOCIOLOGÍA
FACULTAD DE CIENCIAS SOCIALES

Guía N°1

Análisis de Datos Multinivel - SOL3051

Estudiante Andreas Laffert
Profesora Camila Ortiz
Ayudante Andres González

martes 10, septiembre 2024

Table 1: Modelos multinivel para confianza política, ideología política y países con presidencia de izquierda

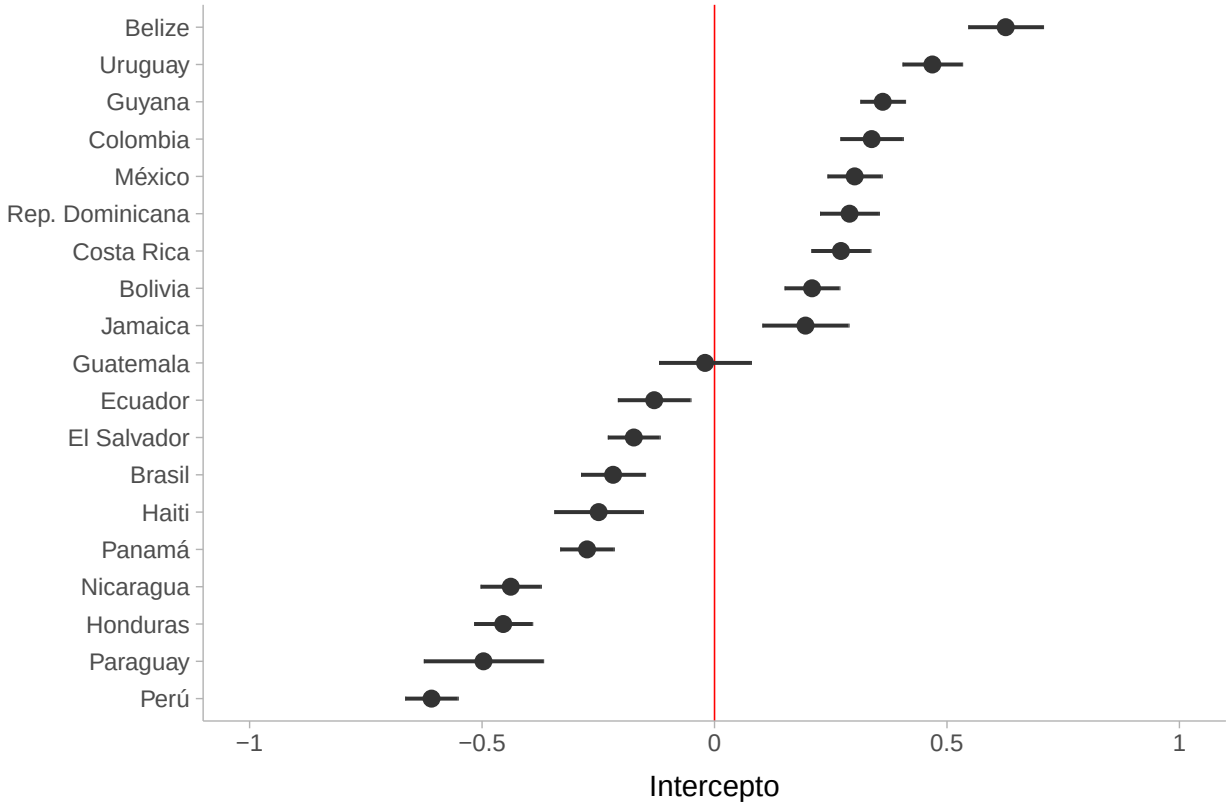
	Modelo 1	Modelo 2	Modelo 3
Intercepto	0.43^{***}	0.34^{**}	0.46^{***}
	(0.09)	(0.12)	(0.13)
Mujer (Ref.= Hombre)	-0.03	-0.03	-0.03
	(0.02)	(0.02)	(0.02)
Edad	0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.01)
Nivel educacional (en años)	-0.01^{***}	-0.01^{**}	-0.01^{**}
	(0.00)	(0.00)	(0.00)
Empleado (Ref.= Desempleado)	-0.06^{**}	-0.06^{***}	-0.06^{***}
	(0.02)	(0.02)	(0.02)
Casado (Ref.= Otro)	0.05^{**}	0.05^{**}	0.05^{**}
	(0.02)	(0.02)	(0.02)
Ideología política	-0.07^{***}	-0.05^{***}	-0.07^{***}
	(0.00)	(0.01)	(0.02)
Presidencia Izquierda (Ref. = Otra)			-0.42
			(0.25)
Ideología política x Presidencia Izquierda (Ref. = Otra)			0.07^{*}
			(0.03)
AIC	52879.32	52533.14	52538.58
BIC	52957.31	52626.73	52647.77
-2*log-likelihood	-26429.66	-26254.57	-26255.29
Num. obs	18012	18012	18012
Num. grupos: Países	19	19	19
Var: Países (Intercepto)	0.14	0.24	0.22
Var: Residual	1.09	1.07	1.07
Var: Países Ideología		0.00	0.00
Cov: Países (Intercepto), Ideología		-0.02	-0.01

Nota: Celdas contienen coeficientes de regresión con errores estándares entre paréntesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Fuente: Elaboración propia en base a LAPOP 2008.

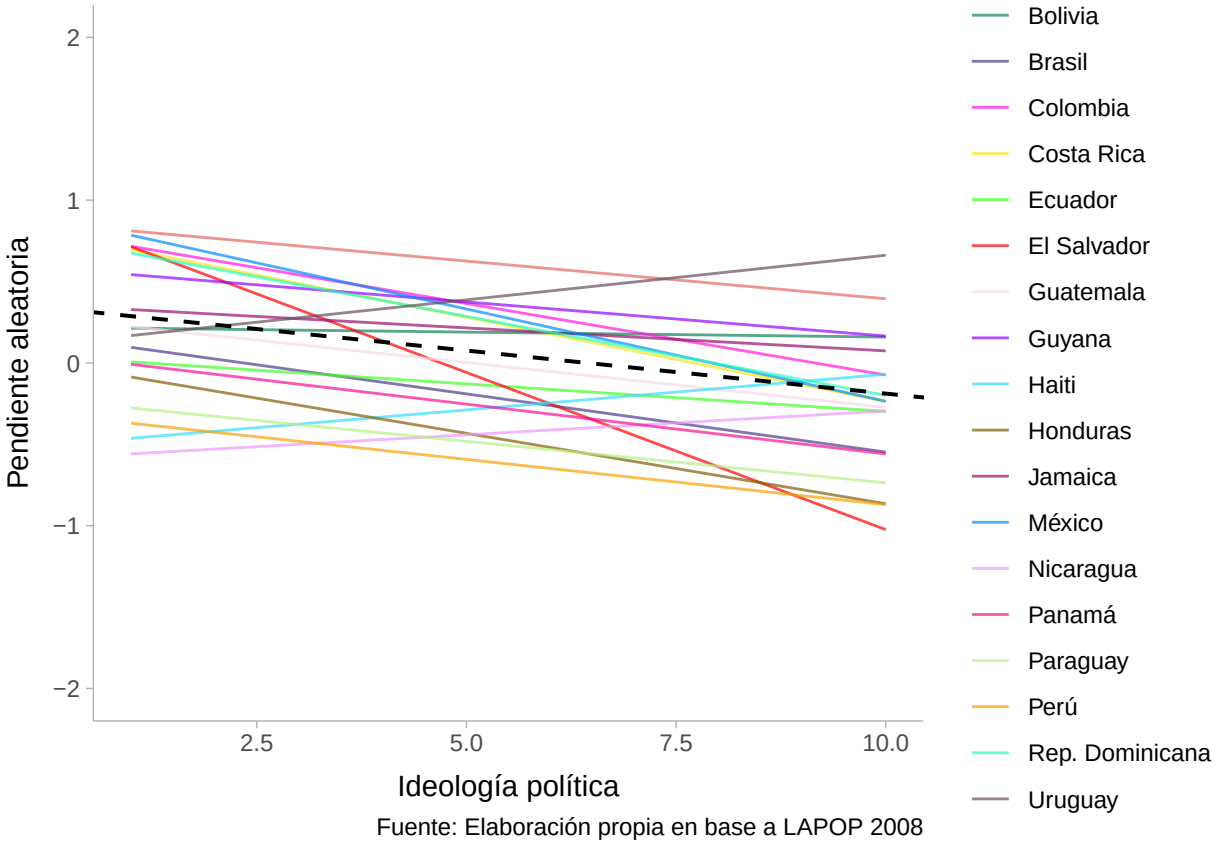
Hola hola

Figure 1: Interceptos aleatorios por país



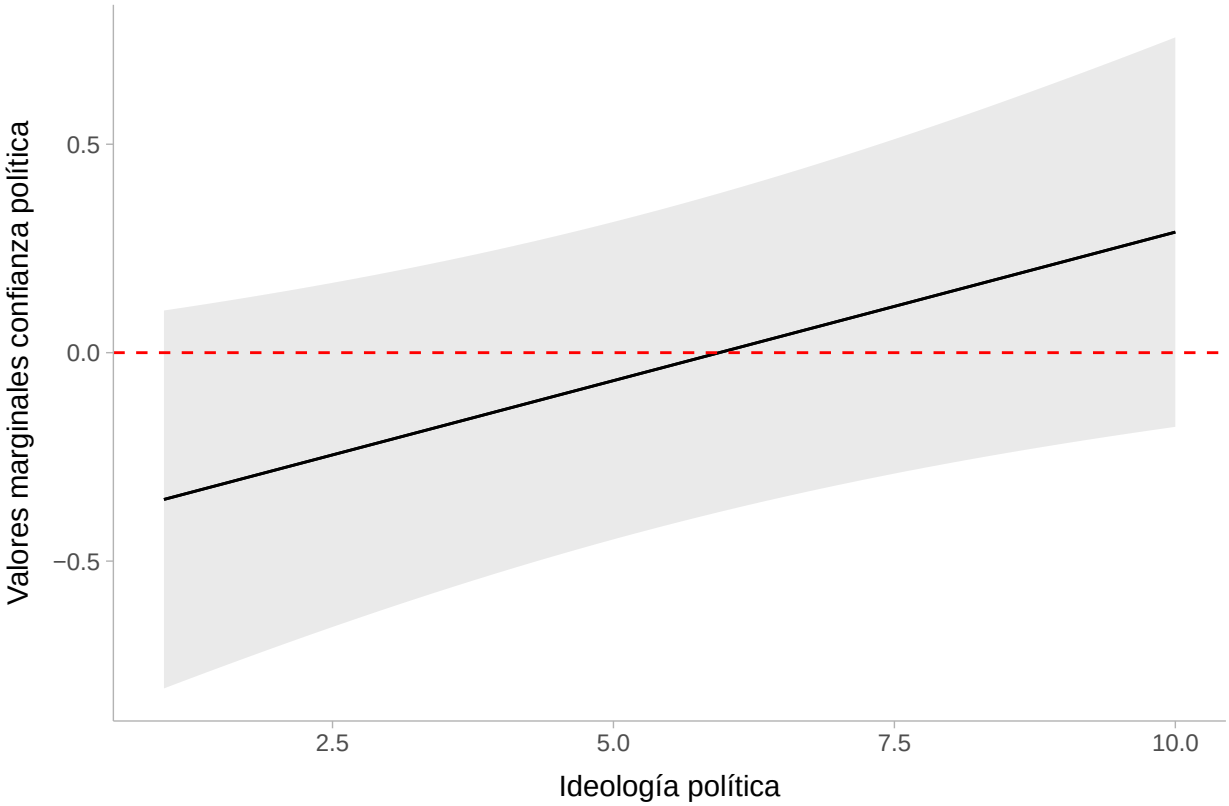
Fuente: Elaboración propia en base a LAPOP 2008

Figure 2: Pendientes aleatorias de ideología política por país



Fuente: Elaboración propia en base a LAPOP 2008

Figure 3: Valores marginales de ideología moderado por presidencia de izquierda



Fuente: Elaboración propia en base a LAPOP 2008

Table 2: Modelos multinivel para confianza política, edad e índice de democracía del país

	Modelo 4	Modelo 5	Modelo 6
Intercepto	0.41^{***}	0.08	0.61^{*}
	(0.08)	(0.41)	(0.24)
Mujer (Ref.= Hombre)	-0.02	-0.02	-0.02
	(0.02)	(0.02)	(0.02)
Edad	0.00	0.00	0.09^{*}
	(0.01)	(0.01)	(0.04)
Nivel educacional (en años)	-0.01^{***}	-0.01^{***}	-0.01^{***}
	(0.00)	(0.00)	(0.00)
Empleado (Ref.= Desempleado)	-0.05^{*}	-0.05^{*}	-0.04^{*}
	(0.02)	(0.02)	(0.02)
Casado (Ref.= Otro)	0.05^{**}	0.05^{**}	0.05^{**}
	(0.02)	(0.02)	(0.02)
Ideología política	-0.07^{***}	-0.07^{***}	-0.07^{***}
	(0.00)	(0.00)	(0.00)
Participación laboral femenina		0.61	
		(0.73)	
Índice Freedom House			-0.08
			(0.09)
Edad x Índice Freedom House			-0.04^{**}
			(0.01)
AIC	52805.69	52805.82	52812.29
BIC	52899.27	52907.20	52921.47
-2*log-likelihood	-26390.84	-26389.91	-26392.14
Num. obs	18012	18012	18012
Num. grupos: Países	19	19	19
Var: Países (Intercepto)	0.10	0.11	0.10
Var: Países Edad	0.00	0.00	0.00
Cov: Países (Intercepto), Edad	0.00	0.00	0.00
Var: Residual	1.09	1.09	1.09

Nota: Celdas contienen coeficientes de regresión con errores estándares entre paréntesis. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Fuente: Elaboración propia en base a LAPOP 2008.

Si bien la primera parte del supuesto de sequential unconfoundedness es más plausible de corregir, la segunda no lo es tanto. El sesgo de variable omitida o de heterogeneidad no observada en estudios observacionales es difícil de combatir, ya que generalmente existen variables no observadas postratamiento que pueden confundir el efecto del mediador en la variable de resultado (Acharya et al., 2016). Precisamente, esto es lo que Pepinsky et al. (2024) cuestionan en la estrategia de identificación de Homola et al. (2024), al asumir que no existen confounders intermedios Z_i no observados que puedan afectar la estimación causal de la distancia a los campos en la intolerancia, lo cual es difícil de sostener.

Table 3: Estadísticos de bondad de ajuste

Modelo	AIC	BIC	Deviance	X2	Df	p-value
Modelo 1	52822.33	52900.32	52802.33			
Modelo 4	52750.41	52844.00	52726.41	75.92	2	< 0.001 ***

Referencias

Acharya, A., Blackwell, M., & Sen, M. (2016). Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects. *American Political Science Review*, *110*(3), 512–529. https://doi.org/10.1017/S0003055416000216

Homola, J., Pereira, M. M., & Tavits, M. (2024). Fixed Effects and Post-Treatment Bias in Legacy Studies. *American Political Science Review*, *118*(1), 537–544. https://doi.org/10.1017/S0003055423001351

Pepinsky, T. B., Goodman, S. W., & Ziller, C. (2024). Causation and History in Legacy Studies: A Reply to Homola, Pereira, and Tavits. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4705690

Código de R

```
knitr::opts_chunk$set(echo = F,
                      warning = F,
                      error = F,
                      message = F)
if (! require("pacman")) install.packages("pacman")

pacman::p_load(tidyverse,
               sjmisc,
               sjPlot,
               lme4,
               easystats,
               influence.ME,
               performance,
               broom.mixed,
               here,
               texreg,
               ggeffects,
               marginaleffects,
               naniar,
               ggdist,
               Polychrome,
               misty,
               kableExtra)

options(scipen=999)
rm(list = ls())

miles <- function(x) {
  format(round(as.numeric(x),0), big.mark = ".")
}

decimales <- function(x) {
  format(round(as.numeric(x), 2), decimal.mark = ",")
}

# set theme

theme_set(theme_ggdist())

options(knitr.kable.NA = "")
```

```
options(knitr.table.format="latex")

load(file = here("input/data/morgan2013.RData"))

names(morgan2013)
glimpse(morgan2013)

# seleccionar ----

db <- morgan2013 %>%
  select(country, ID, trustgov, sex = female, age, educ, employed, marrie
         race, ideology = left, leftpres, FLP, fhouse) %>%
  sjlabelled::remove_all_labels() %>%
  janitor::clean_names() %>%
  as_tibble()

# filtrar: no -----

# recodificar y transformar ----

# trust
sjmisc::descr(db$in_trust)

# sexo
frq(db$sex)

db$sex <- car::recode(db$sex,
                      recodes= c("0='Hombre';1='Mujer'"),
                      levels = c("Hombre","Mujer"),
                      as.factor = T)

# edad
sjmisc::descr(db$age)
frq(db$age)

db$age_f <- car::recode(db$age,
                      recodes = c("1='Tramo 1';
                                  2='Tramo 2';
                                  3='Tramo 3';
                                  4='Tramo 4';
                                  5='Tramo 5';
```

```
                                    6='Tramo 6'"),
                        levels = c("Tramo 1",
                                   "Tramo 2",
                                   "Tramo 3",
                                   "Tramo 4",
                                   "Tramo 5",
                                   "Tramo 6"),
                        as.factor = T
)

# educ
sjmisc::descr(db$educ)

# employed
frq(db$employed)

db$employed <- car::recode(db$employed,
                        recodes= c("0='Desempleado';1='Empleado'"),
                        levels = c("Desempleado","Empleado"),
                        as.factor = T)
# married
frq(db$married)

db$married <- car::recode(db$married,
                        recodes= c("0='No';1='Sí'"),
                        levels = c("No","Sí"),
                        as.factor = T)

# race
frq(db$race)

db$race <- car::recode(db$race,
                        recodes= c("0='Otro';1='Blanco'"),
                        levels = c("Otro","Blanco"),
                        as.factor = T)

# ideology
frq(db$ideology)

# left
frq(db$leftpres)
```

```r
db$leftpres <- car::recode(db$leftpres,
                      recodes= c("0='No';1='Sí'"),
                      levels = c("No","Sí"),
                      as.factor = T)


# flp
sjmisc::descr(db$flp)

# fhouse
sjmisc::descr(db$fhouse)

# id
sjmisc::descr(db$id)

# country
frq(db$country)

# casos perdidos -----

colSums(is.na(db))

n_miss(db)

prop_miss(db)*100

miss_var_summary(db)

miss_var_table(db)

vis_miss(db) + theme(axis.text.x = element_text(angle=80))

db <- na.omit(db)

# Null model
model_0 <- lmer(in_trust ~ 1 + (1 | country),
                data = db, REML = T)

performance::icc(model_0, by_group = T)
## ICC Country = 0.11
```

```r
# Influence test
inf_m0 <- influence(model_0, group = "country")

# D cook
cooks.distance(inf_m0, parameters = 1, sort = T) # cut point is 4/19

n_country <- length(unique(db$country))

plot(inf_m0, which="cook",
     cutoff=(4/n_country), sort=TRUE,
     xlab="Distancia de Cook",
     ylab="País", width=60, height=40)

# no obs influyentes

# Modelo 1: Indicadores N1
model_1 <- lmer(in_trust ~ 1 + sex + age + educ + employed + married +
                race + ideology + (1 | country),
                data = db,
                REML = T)

# Modelo 2: Pendiente aleatoria ideology
model_2 <- lmer(in_trust ~ 1 + sex + age + educ + employed + married +
                race + ideology + (1 + ideology| country),
                data = db,
                REML = T)

# Modelo 3: Interaccion ideology y leftpres
model_3 <- lmer(in_trust ~ 1 + sex + age + educ + employed + married +
                race + ideology + leftpres + ideology*leftpres +
                (1 + ideology| country),
                data = db,
                REML = T)

# Modelo 4: Pendiente aleatoria edad
model_4 <- lmer(in_trust ~ 1 + sex + age + educ + employed + married +
                race + ideology + (1 + age| country),
                data = db,
                REML = T)

# Modelo 5: Pendiente aleatoria edad + flp
```

```r
model_5 <- lmer(in_trust ~ 1 + sex + age + educ + employed + married +
                race + ideology + flp + (1 + age| country),
                data = db,
                REML = T)

# Modelo 6: Pendiente aleatoria edad + flp
model_6 <- lmer(in_trust ~ 1 + sex + age + educ + employed + married +
                race + ideology + age*fhouse + (1 + age| country),
                data = db,
                REML = T)


ccoef <- list(
  "(Intercept)" = "Intercepto",
  sexMujer = "Mujer (Ref.= Hombre)",
  age = "Edad",
  educ = "Nivel educacional (en años)",
  employedEmpleado = "Empleado (Ref.= Desempleado)",
  marriedSí = "Casado (Ref.= Otro)",
  race = "Blanco (Ref.= Otro)",
  ideology = "Ideología política",
  leftpresSí = "Presidencia Izquierda (Ref. = Otra)",
  "ideology:leftpresSí" = "Ideología política x Presidencia Izquierda (Re


texreg::texreg(list(model_1, model_2, model_3),
               custom.model.names = c("Modelo 1",
                                      "Modelo 2",
                                      "Modelo 3"),
               caption = paste("(\\#tab:table1)","Modelos multinivel para
               stars = c(0.05, 0.01, 0.001),
               custom.coef.map = ccoef,
               custom.note = "\\item Nota: Celdas contienen coeficientes
               threeparttable = T,
               leading.zero = T,
               float.pos = "h!",
               use.packages = F,
               booktabs = TRUE,
               scalebox = 0.9,
               custom.gof.names = c("AIC",
                                    "BIC",
```

```
                                        "-2*log-likelihood",
                                        "Num. obs",
                                        "Num. grupos: Países",
                                        "Var: Países (Intercepto)",
                                        "Var: Residual",
                                        "Var: Países Ideología",
                                        "Cov: Países (Intercepto), Ideología"
                                        ))
sjPlot::plot_model(model_1,
                   type = "re",
                   vline.color = "red",
                   grid = F,
                   sort.est = "sort.all",
                   ci.lvl = .95,
                   colors = "grey20") +
  labs(title = NULL,
       y = "Intercepto",
       caption = "Fuente: Elaboración propia en base a LAPOP 2008")+
  theme_ggdist()


stats_m2 <- broom.mixed::tidy(model_2)

ideology_rang <- seq(1:10)
in_trust2 <- as.data.frame(sapply(ideology_rang,function(x)fixef(model_2)
                                        fixef(model_2)["sexMujer"]*0 +
                                        fixef(model_2)["age"]*2.74 +
                                        fixef(model_2)["educ"]*9.23 +
                                        fixef(model_2)["employedEmpleado"]*1
                                        fixef(model_2)["marriedSí"]*1 +
                                        fixef(model_2)["raceBlanco"]*0 +
                                        fixef(model_2)["ideology"]*x +
                                        ranef(model_2)$country[,1] +
                                        ranef(model_2)$country[,2]*x))

in_trust2$country <- seq(1:19)
in_trust2 <- reshape(in_trust2,
                     direction = "long", #formato long
                     v.names = "pred",
                     varying = list(names(in_trust2)[1:10]),
                     idvar = c("country"),
```

```
                      timevar = "ideology")

colores <- setNames(createPalette(19, c("#E16462", "#177b4b", "#0D0887")))


in_trust2 %>%
  mutate(country = factor(country,
                          levels = 1:19,
                          labels = levels(db$country))) %>%
  ggplot(aes(x = ideology, y = pred, group = country, color = country)) +
  geom_line(alpha = 0.7) +
  ylim(-2,2) +
  geom_abline(intercept = stats_m2$estimate[stats_m2$term == "(Intercept)
              slope = stats_m2$estimate[stats_m2$term == "ideology"],
              color = "black",
              linetype = "dashed",
              linewidth = 0.7) +
  scale_color_manual(values = colores) +
  labs(y = "Pendiente aleatoria",
       x = "Ideología política",
       color = "País",
       caption = "Fuente: Elaboración propia en base a LAPOP 2008")


# manual

varcomp_0=as.data.frame(VarCorr(model_0))
tau00_0=varcomp_0[1,4]
sigma2_0=varcomp_0[2,4]

varcomp_1=as.data.frame(VarCorr(model_1))
tau00_1=varcomp_1[1,4]
sigma2_1=varcomp_1[2,4]

R2_1_L1=(sigma2_0-sigma2_1)/sigma2_0
R2_1_L1

R2_2_L1=(tau00_0-tau00_1)/tau00_0
R2_2_L1

# misty
```

```r
db_n <- db %>%
  mutate(
    across(.cols = c(sex, age, employed, married, race),
           .fns = ~ as.numeric(.))
  )


model_1_n <- lmer(in_trust ~ 1 + sex + age + educ + employed + married +
                  race + ideology + (1 | country),
                  data = db_n,
                  REML = T)

misty::multilevel.r2(model = model_1_n, print = "all")


plot_slopes(model_3,
            variables = "leftpres",
            condition = "ideology",
            conf_level = .95) +
  geom_hline(yintercept = 0,
             color = "red",
             linetype = "dashed") +
  labs(y = "Valores marginales confianza política",
       x = "Ideología política",
       caption = "Fuente: Elaboración propia en base a LAPOP 2008",
       title = NULL)


ccoef <- list(
  "(Intercept)" = "Intercepto",
  sexMujer = "Mujer (Ref.= Hombre)",
  age = "Edad",
  educ = "Nivel educacional (en años)",
  employedEmpleado = "Empleado (Ref.= Desempleado)",
  marriedSí = "Casado (Ref.= Otro)",
  race = "Blanco (Ref.= Otro)",
  ideology = "Ideología política",
  flp = "Participación laboral femenina",
  fhouse = "Índice Freedom House",
  "age:fhouse" = "Edad x Índice Freedom House"
)
```

```
texreg::texreg(list(model_4, model_5, model_6),
               custom.model.names = c("Modelo 4",
                                      "Modelo 5",
                                      "Modelo 6"),
               caption = paste("(\\#tab:table2)","Modelos multinivel para
               stars = c(0.05, 0.01, 0.001),
               custom.coef.map = ccoef,
               custom.note = "\\item Nota: Celdas contienen coeficientes
               threeparttable = T,
               leading.zero = T,
               float.pos = "h!",
               use.packages = F,
               booktabs = TRUE,
               scalebox = 0.9,
               custom.gof.names = c("AIC",
                                    "BIC",
                                    "-2*log-likelihood",
                                    "Num. obs",
                                    "Num. grupos: Países",
                                    "Var: Países (Intercepto)",
                                    "Var: Países Edad",
                                    "Cov: Países (Intercepto), Edad",
                                    "Var: Residual"

                                    ))

#performance::test_likelihoodratio(model_1, model_4)

res_fit1 <- anova(model_1, model_4)

fit_tab <- res_fit1[c(2,3,5,6,7,8)] %>% as.tibble(.)

fit_tab$mod <- c("Modelo 1", "Modelo 4")

fit_tab$p <- if_else(fit_tab$`Pr(>Chisq)` < 0.001, "< 0.001 ***", NA)

fit_tab$Chisq <- round(fit_tab$Chisq, digits = 2)

fit_tab <- fit_tab %>%
  select(mod, AIC, BIC, deviance, Chisq, Df, p)
```

```
colnames <- c("Modelo", "AIC", "BIC", "Deviance", "X2", "Df", "p-value")

fit_tab <- kableExtra::kable(fit_tab,
                             format = "latex",
                             col.names = colnames,
                             row.names = F,
                             booktabs = T,
                             caption = paste("(\\#tab:table3)","Estadísti
  kableExtra::kable_styling(latex_options = "hold_position",
                            position = "center") %>%
  kableExtra::kable_styling(bootstrap_options = c("striped", "hover", "co
  column_spec(1, width = "2cm") %>%
  row_spec(0, bold = T)

fit_tab
```