

# Laboratorio 3

## Datos politicos de Uruguay

Nicolás Schmidt – Fabricio Carneiro

UMAD – FCS

2020-10-13

**¿A medida que aumenta la volatilidad electoral aumenta la fragmentacion del sistema de partidos?**

# Variables

- **Variable dependiente**

Fragmentacion  $\rightarrow$  Número Efectivo de Partidos =  $\frac{1}{\sum_{i=1}^n p_i^2}$

- **Variable independiente**

Volatilidad electoral  $\rightarrow$  Pedersen =  $\frac{1}{2} \sum_{i=1}^n |P_i \nabla|$

# Modelo

- Formalizado

$$\text{enp} = \alpha + \beta_1(\text{eVolat}) + \epsilon$$

- En R

```
modelo <- lm(nep ~ evolat, data = datos)
```

# Ingredientes

Necesitamos dos variables (nep y evolat).

Ambas variables son indicadores que requieren tener resultados electorales.

Vamos a necesitar dos paquetes que nos proporcionan datos electorales de uruguay (`{Boreluy}`) y funcioanes que permiten calcular esos indicadores caracteristicos de los sistemas de partidos (`{esaps}`).

**Cargamos los paquetes:**

```
library(Boreluy)  # datos electorales de UY  
library(esaps)    # indicadores de SP y SE
```

# DATA! Paso 1: datos electorales

```
datos <-  
  elecciones("Presidencial") %>%  
  [.>=1942] %>%  
  purrr::map_df(nacional_uy) %>%  
  as_esaps()
```

```
## # A tibble: 158 x 5
```

	election	unit	party	votes	seats
	<chr>	<chr>	<chr>	<dbl>	<dbl>
## 1	1942	Uruguay	Partido Colorado	57.2	58.6
## 2	1942	Uruguay	Partido Nacional	22.8	23.2
## 3	1942	Uruguay	Partido Nacional Independiente	11.7	11.1
## 4	1942	Uruguay	Union Civica	4.25	4.04
## 5	1942	Uruguay	Partido Comunista del Uruguay	2.49	2.02
## 6	1942	Uruguay	Partido Socialista	1.57	1.01
## 7	1942	Uruguay	Partido la Concordancia	0.00696	0
## 8	1946	Uruguay	Partido Colorado	46.3	47.5
## 9	1946	Uruguay	Partido Nacional	31.1	31.3
## 10	1946	Uruguay	Partido Nacional Independiente	9.39	9.09
## #	... with 148 more rows				

# DATA! Paso 2: Indicadores

## Número efectivo de partidos

```
enp <- esaps::enp(datos[, -5])  
print(enp)
```

##	<i>election</i>	<i>unit</i>	<i>enp</i>
## 1	1942	Uruguay	2.53
## 2	1946	Uruguay	3.07
## 3	1950	Uruguay	2.65
## 4	1954	Uruguay	2.60
## 5	1958	Uruguay	2.55
## 6	1962	Uruguay	2.40
## 7	1966	Uruguay	2.44
## 8	1971	Uruguay	2.91
## 9	1984	Uruguay	3.09
## 10	1989	Uruguay	3.66
## 11	1994	Uruguay	3.67
## 12	1999	Uruguay	3.29
## 13	2004	Uruguay	2.61
## 14	2009	Uruguay	2.86
## 15	2014	Uruguay	2.92
## 16	2019	Uruguay	3.76

# DATA! Paso 2: Indicadores

## Volatilidad electoral

```
evolat <- esaps::evolat(datos, method = "Pedersen")  
print(evolat)
```

##	<i>election</i>	<i>unit</i>	<i>eVolat</i>
## 1	1946	Uruguay	13.13
## 2	1950	Uruguay	6.02
## 3	1954	Uruguay	6.33
## 4	1958	Uruguay	17.83
## 5	1962	Uruguay	7.69
## 6	1966	Uruguay	14.20
## 7	1971	Uruguay	21.15
## 8	1984	Uruguay	7.01
## 9	1989	Uruguay	14.14
## 10	1994	Uruguay	11.59
## 11	1999	Uruguay	11.10
## 12	2004	Uruguay	26.18
## 13	2009	Uruguay	8.18
## 14	2014	Uruguay	5.50
## 15	2019	Uruguay	15.13

# DATA! Paso 3: merge

```
datos <- right_join(enp, evolat, by = c('election', 'unit'))  
print(datos)
```

```
##      election      unit  enp  eVolat  
## 1      1946 Uruguay 3.07  13.13  
## 2      1950 Uruguay 2.65   6.02  
## 3      1954 Uruguay 2.60   6.33  
## 4      1958 Uruguay 2.55  17.83  
## 5      1962 Uruguay 2.40   7.69  
## 6      1966 Uruguay 2.44  14.20  
## 7      1971 Uruguay 2.91  21.15  
## 8      1984 Uruguay 3.09   7.01  
## 9      1989 Uruguay 3.66  14.14  
## 10     1994 Uruguay 3.67  11.59  
## 11     1999 Uruguay 3.29  11.10  
## 12     2004 Uruguay 2.61  26.18  
## 13     2009 Uruguay 2.86   8.18  
## 14     2014 Uruguay 2.92   5.50  
## 15     2019 Uruguay 3.76  15.13
```



# DATA! Resumen (solo 8 lineas de código!)

```
## PASO 1
1 datos <-
2 elecciones("Presidencial") %>%
3 .[.>=1942] %>%
4 purrr::map_df(nacional_uy) %>%
5 as_esaps()

## PASO 2
6 enp <- esaps::enp(datos[, -5])
7 evolat <- esaps::evolat(datos, method = "Pedersen")

## PASO 3
8 datos <- right_join(enp, evolat, by = c('election', 'unit'))
```

# Explorando datos

Para graficar las dos series de nuestras dos variables necesitamos pasar de wide a long los datos.

```
print(datos)
##      election      unit  enp eVolat
## 1      1946 Uruguay 3.07  13.13
## 2      1950 Uruguay 2.65   6.02
## 3      1954 Uruguay 2.60   6.33
## 4      1958 Uruguay 2.55  17.83
## 5      1962 Uruguay 2.40   7.69
## 6      1966 Uruguay 2.44  14.20
## 7      1971 Uruguay 2.91  21.15
## 8      1984 Uruguay 3.09   7.01
## 9      1989 Uruguay 3.66  14.14
## 10     1994 Uruguay 3.67  11.59
## 11     1999 Uruguay 3.29  11.10
## 12     2004 Uruguay 2.61  26.18
## 13     2009 Uruguay 2.86   8.18
## 14     2014 Uruguay 2.92   5.50
## 15     2019 Uruguay 3.76  15.13
```

# Explorando datos: pivoteando

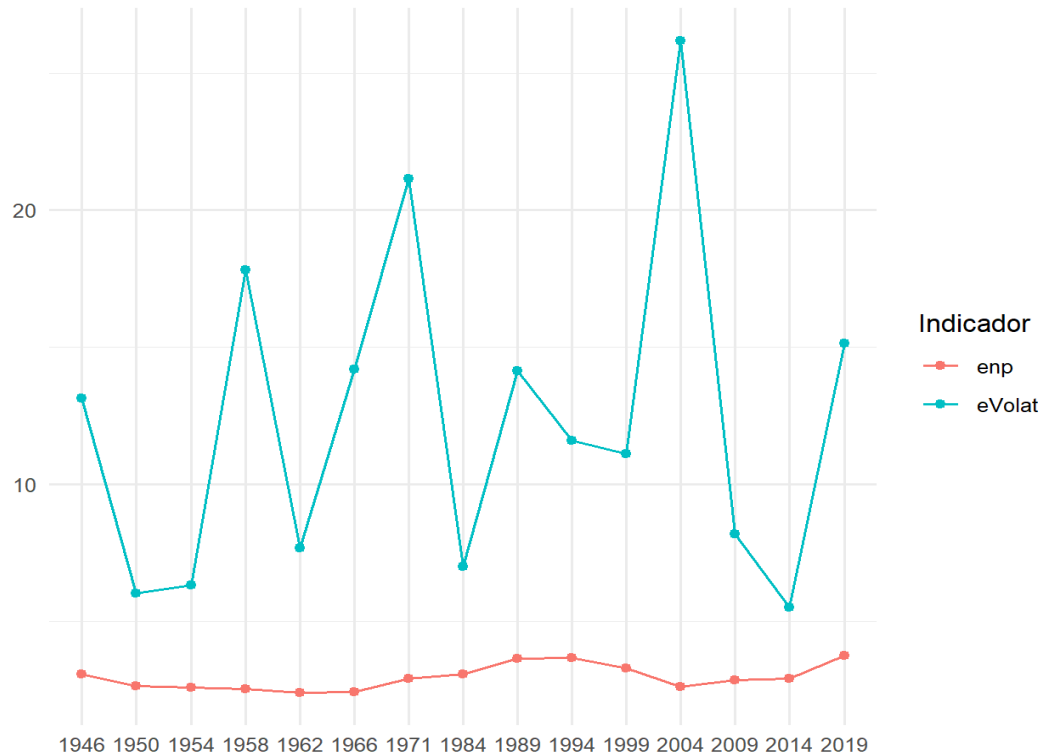
```
data_long <- tidyr::pivot_longer(data = datos,  
                                cols = c("enp", "eVolat"),  
                                names_to = "Indicador",  
                                values_to = "valor")
```

```
data_long
```

```
## # A tibble: 30 x 4  
##   election unit   Indicador valor  
##   <chr>      <chr>   <chr>      <dbl>  
## 1 1946      Uruguay enp         3.07  
## 2 1946      Uruguay eVolat    13.1  
## 3 1950      Uruguay enp         2.65  
## 4 1950      Uruguay eVolat    6.02  
## 5 1954      Uruguay enp         2.6  
## 6 1954      Uruguay eVolat    6.33  
## 7 1958      Uruguay enp         2.55  
## 8 1958      Uruguay eVolat    17.8  
## 9 1962      Uruguay enp         2.4  
## 10 1962     Uruguay eVolat    7.69  
## # ... with 20 more rows
```

# Explorando datos: graficando

```
ggplot(data_long, aes(x = election, y = valor,  
                      group = Indicador, color = Indicador)) +  
  geom_line() + geom_point() + theme_minimal() + labs(x = "", y = "")
```



# Explorando datos: resumen

```
# nombres de las variables
names(datos)
## [1] "election" "unit"      "enp"      "eVolat"

# resumen de las variables de interés
summary(datos[, 3:4])
##           enp           eVolat
##  Min.      :2.400   Min.      : 5.50
## 1st Qu.:2.605   1st Qu.: 7.35
##  Median :2.910   Median :11.59
##   Mean   :2.965   Mean    :12.35
## 3rd Qu.:3.190   3rd Qu.:14.66
##   Max.   :3.760   Max.    :26.18
```

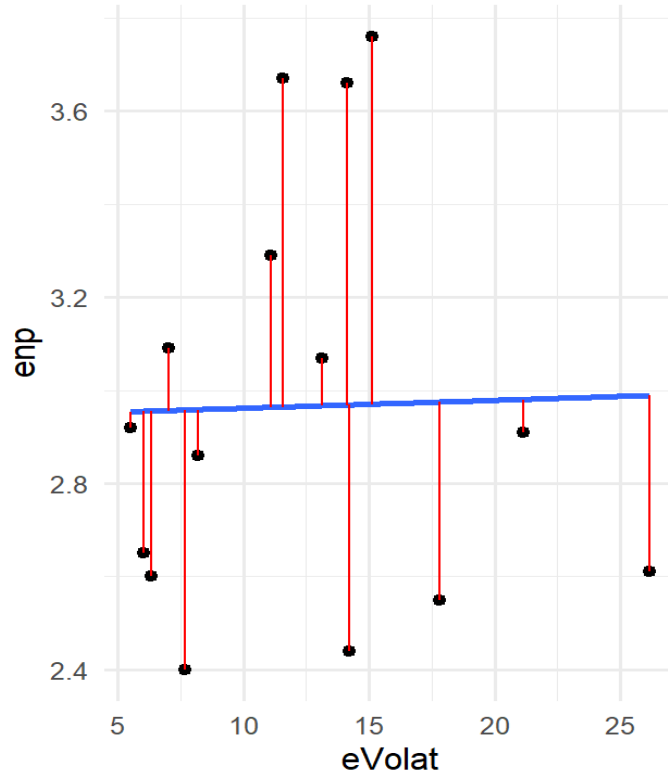
# Modelo!

```
modelo <- lm(enp ~ eVolat, data = datos)
summary(modelo)
##
## Call:
## lm(formula = enp ~ eVolat, data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.55745 -0.36695 -0.07024  0.23024  0.78995
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.944437   0.285540  10.312 1.26e-07 ***
## eVolat       0.001693   0.020926   0.081   0.937
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.471 on 13 degrees of freedom
## Multiple R-squared:  0.000503,    Adjusted R-squared:  -0.07638
## F-statistic: 0.006543 on 1 and 13 DF,  p-value: 0.9368
```

# Modelo! lectura paso a paso

1 - [Call] Formula utilizada para ajustar un modelo lineal con los datos.

2 - [Residuals]



La distribución de los residuos. Los residuos son la distancia entre los valores observados (puntos negros) y los valores predichos por el modelo (recta azul). En el grafico son las rectas rojas. ¡¡Son distancias verticales no horizontales!!

La media de los residuos es 0.

```
mean(modelo$residuals)
## [1] 5.540273e-18
```

Se deber observar la simetría de la distribución, por supuesto la distribución de los errores sigue una distribución gaussiana.

# Modelo! lectura paso a paso

## 3 - [Coefficients]

- **Estimate** Intercepto es el valor esperado de  $y$  cuando  $x = 0$ . En este caso cuando a un nivel 0 de volatilidad hay un número efectivo de partidos esperado de casi 3. El estimador de `evol` nos informa la dirección y la pendiente de la recta de regresión. Se debe leer de la siguiente manera: es el cambio medio esperado en  $y$  ante el incremento de una unidad de  $x$  (todo lo demás constante)
- **Std. Error** Es el error estándar de la estimación.

```
confint(modelo)
##                2.5 %      97.5 %
## (Intercept)  2.32756526 3.56130791
## eVolat      -0.04351581 0.04690118
```

- **t value** Prueba T para determinar si el coeficiente es significativamente distinto de 0.
- **Pr(> |t|)** Valor  $p$  para la prueba de hipótesis T.



# Modelo! lectura paso a paso

## Análisis de Varianza:

Determinar la variabilidad de los datos que es explicada por el modelo y la variabilidad que no es explicada (residuos). Esto requiere una prueba de significancia para ver si la variabilidad explicada es suficientemente grande.

```
anova(modelo)
## Analysis of Variance Table
##
## Response: enp
##           Df Sum Sq Mean Sq F value Pr(>F)
## eVolat      1 0.00145  0.001452   0.0065 0.9368
## Residuals 13 2.88452  0.221886
```

$$SST = \sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2$$

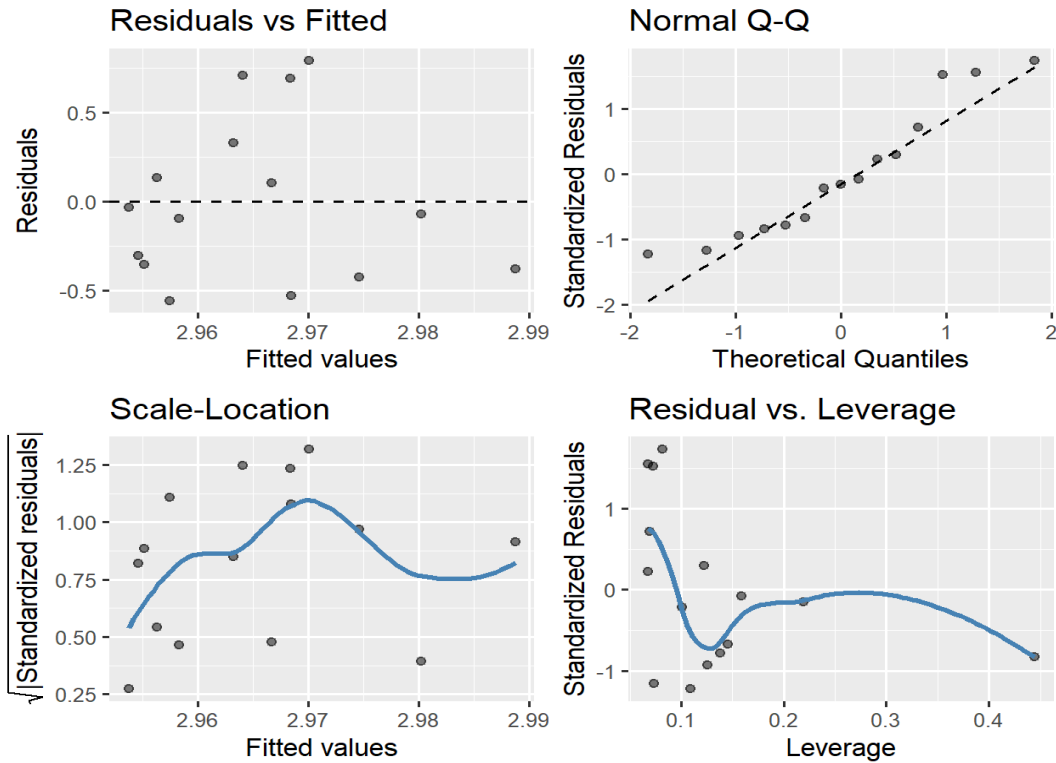
Variabilidad total = Variabilidad residual + Variabilidad recta

# Modelo! supuestos sobre los errores

- Normalidad:  $\epsilon_i \sim \mathcal{N}(\mu, \sigma^2)$
- Media cero:  $E(\epsilon_i) = 0$
- Homocedasticidad:  $V(e_i) = \sigma^2$
- Incorrelacion:  $Cov(\epsilon_i, \epsilon_j) = 0$

# Modelo! Visualización

```
gglm(modelo)
```



# Modelo!: yapa!

```
report(modelo)
```

We fitted a linear model (estimated using OLS) to predict enp with eVolat (formula = enp ~ eVolat). Standardized parameters were obtained by fitting the model on a standardized version of the dataset. Effect sizes were labelled following Cohen's (1988) recommendations.

The model explains a not significant and very weak proportion of variance ( $R^2 = 5.03e-04$ ,  $F(1, 13) = 6.54e-03$ ,  $p = 0.937$ , adj.  $R^2 = -0.08$ ). The model's intercept, corresponding to enp = 0 and eVolat = 0, is at 2.94 (SE = 0.29, 95% CI [2.33, 3.56],  $p < .001$ ). Within this model:

- The effect of eVolat is positive and can be considered as very small and not significant (beta =  $1.69e-03$ , SE = 0.02, 95% CI [-0.04, 0.05], std. beta = 0.02,  $p = 0.937$ ).

# {speech}

```
library(speech)
text <- speech_build("hablamanini.pdf", compiler = TRUE)
text
```

## # A tibble: 10 x 6

##	legislator	legislature	chamber	date	id	speech
##	<chr>	<dbl>	<chr>	<date>	<chr>	<chr>
##	1 ABREU	49	CAMARA DE~	2020-09-09	habla~	SEÑOR ABRI
##	2 ASIAIN	49	CAMARA DE~	2020-09-09	habla~	SEÑORA ASI
##	3 BIANCHI	49	CAMARA DE~	2020-09-09	habla~	SEÑORA BI
##	4 CAMY	49	CAMARA DE~	2020-09-09	habla~	SEÑOR CAM
##	5 CARRERA	49	CAMARA DE~	2020-09-09	habla~	SEÑOR CARI
##	6 LAZO	49	CAMARA DE~	2020-09-09	habla~	SEÑORA LA
##	7 MANINI RIOS	49	CAMARA DE~	2020-09-09	habla~	SEÑOR MANI
##	8 PONTE	49	CAMARA DE~	2020-09-09	habla~	SEÑORA POI
##	9 SANGUINETTI	49	CAMARA DE~	2020-09-09	habla~	SEÑORA SAI
##	10 TOPOLANSKY	49	CAMARA DE~	2020-09-09	habla~	SEÑORA TOI

# {speech}

```
minchar <- function(string, min = 3){  
  string <- stringr::str_remove_all(string, "[[:punct:]]")  
  string <- unlist(strsplit(string, " "))  
  string[nchar(string) > min]  
}  
  
nube <- function(datos, min.char = 4, rm.palabras = c("señor", "señorita",  
  datos$speech %>%  
  minchar(., min = min.char) %>%  
  quanteda::corpus() %>%  
  quanteda::dfm(remove = rm.palabras) %>%  
  quanteda::textplot_wordcloud(color = rev(RColorBrewer::brewer.pal(12, "RdYlGn"))))  
}
```

# {speech}

```
nube(datos = subset(text, legislator == "MANINI RIOS"), min.char = 3)
```



A word cloud visualization showing the frequency of words in the speech of Manini Ríos. The words are arranged in a cluster, with 'artigas' being the most prominent word in the center. Other visible words include 'políticas', 'pueblos', 'derrota', 'años', 'eran', 'paraguay', and 'solo'.

paraguay  
solo políticas  
eran  
artigas  
pueblos  
derrota años