

Análisis de Datos Panel

Derek Martell

2025-06-03

Introducción

En este análisis se utiliza un modelo de datos panel para estudiar los determinantes del PIB per cápita en una muestra de países. Usaremos tres tipos de modelos: Pooled OLS, Efectos Fijos y Efectos Aleatorios, y luego evaluaremos cuál se ajusta mejor a los datos. # Carga de librerías

```
library(plm)      # Modelos de datos panel
library(readxl)   # Para lectura de archivos Excel (no usado aquí, pero puede ser útil)
```

Importación de datos

```
# Ruta del archivo CSV
path <- "C:/eco-poli/TRABAJO FINAL/CÓDIGO EN R/merged_df_sin_ecuador.csv"
df <- read.csv(path)
```

Conversión a estructura panel

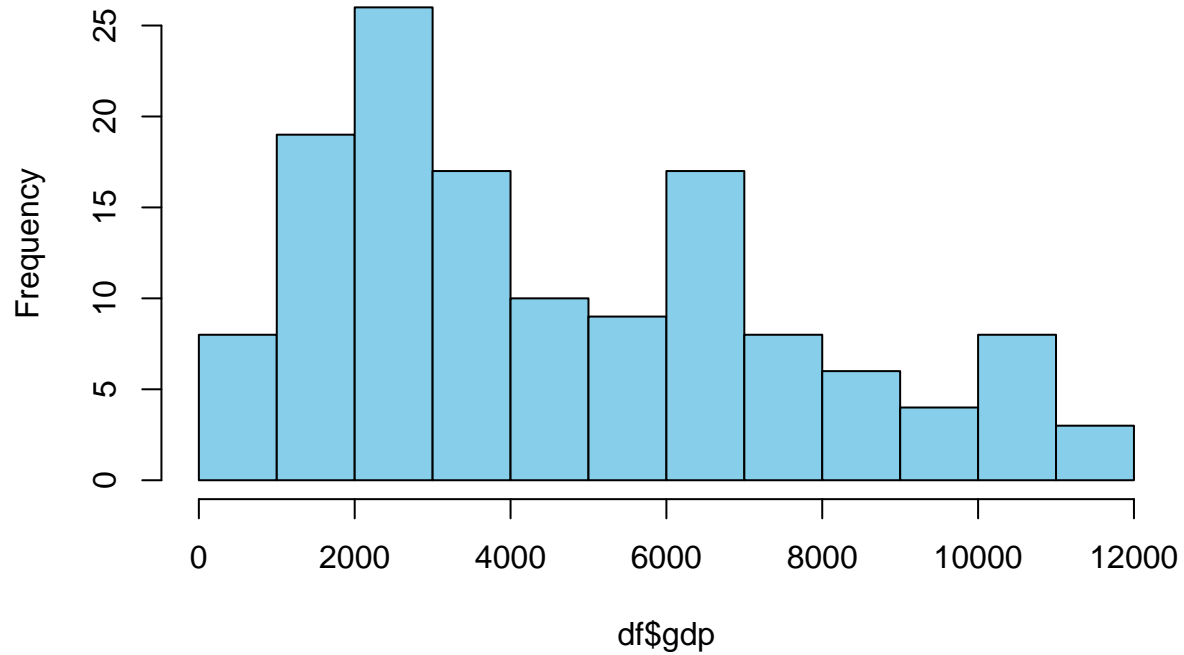
```
df <- pdata.frame(df, index = c("Country", "Index.Year"))
```

- Se especifica que los datos están organizados por país y año.

Análisis exploratorio

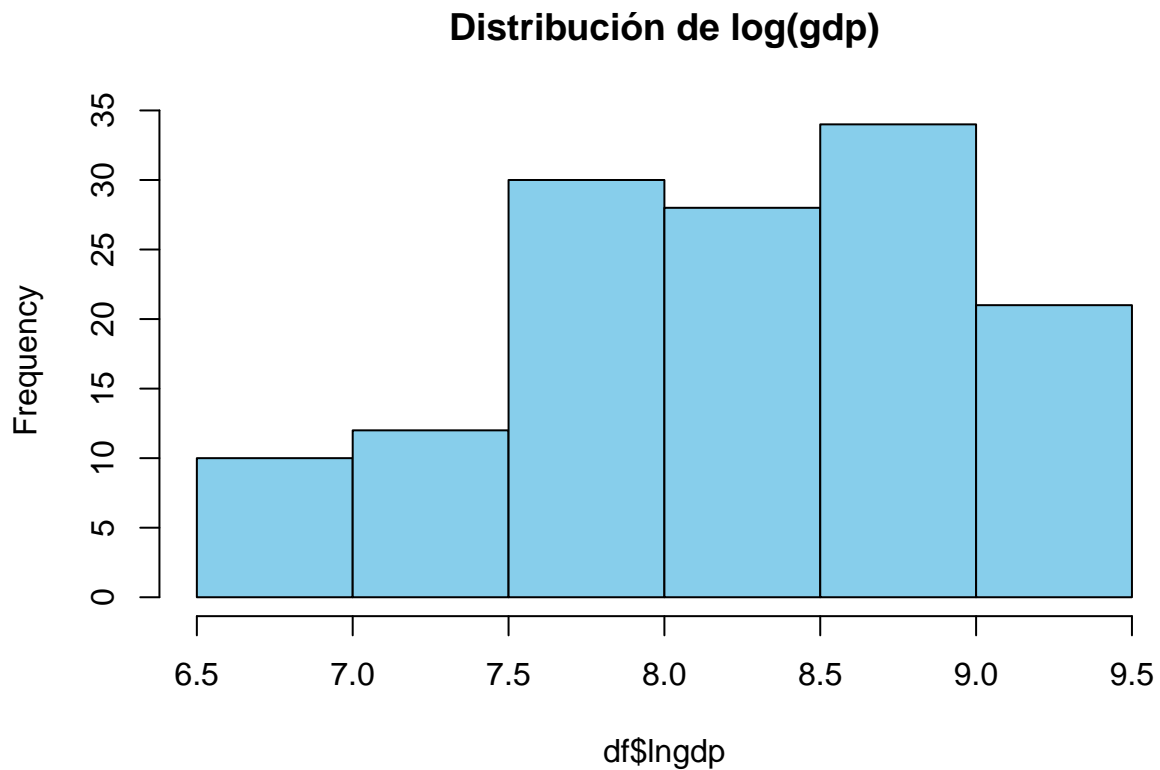
```
hist(df$gdp, main = "Distribución de gdp", col = "skyblue")
```

Distribución de gdp



- El histograma muestra la distribución del PIB per cápita. Suele ser sesgada, por lo que aplicamos una transformación logarítmica.

```
df$lngdp <- log(df$gdp)
hist(df$lngdp, main = "Distribución de log(gdp)", col = "skyblue")
```



- El logaritmo ayuda a normalizar la distribución.

Correlación entre variables explicativas

```
cor(df[, c("capi", "art", "labor", "Overall.Score")])
```

```
##           capi      art      labor Overall.Score
## capi      1.0000000 0.4480974 0.5426616    0.2566290
## art       0.4480974 1.0000000 0.9134098    0.2962500
## labor     0.5426616 0.9134098 1.0000000    0.3391569
## Overall.Score 0.2566290 0.2962500 0.3391569    1.0000000
```

- Se evalúa multicolinealidad. Correlaciones altas podrían afectar la estimación.

VIF

```
library(car)
```

```
## Loading required package: carData
```

```
# Creamos un modelo lineal simple (sin estructura de panel) solo para evaluar multicolinealidad
lm_model <- lm(gdp ~ capi + art + labor + Overall.Score, data = df)

# Calculamos el VIF
vif(lm_model)
```

```
##           capi           art           labor Overall.Score
##      1.456625      6.158046      7.099257      1.140191
```

Modelo Pooled (OLS)

```
reg_pool <- plm(lngdp ~ capi + art + labor + Overall.Score, data = df, model = "pooling")
summary(reg_pool)
```

```
## Pooling Model
##
## Call:
## plm(formula = lngdp ~ capi + art + labor + Overall.Score, data = df,
##      model = "pooling")
##
## Balanced Panel: n = 5, T = 27, N = 135
##
## Residuals:
##      Min.    1st Qu.    Median    3rd Qu.    Max.
## -1.015091 -0.260130 -0.027057  0.304282  0.843768
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## (Intercept)  6.9665e+00  4.0327e-01 17.2750 < 2.2e-16 ***
## capi         5.2774e-02  1.3149e-02  4.0137 0.0001004 ***
## art        -6.0900e-06  1.6913e-05 -0.3601 0.7193824
## labor        3.1547e-08  5.8681e-09  5.3760 3.416e-07 ***
## Overall.Score -5.1875e-03  5.8517e-03 -0.8865 0.3769895
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    65.451
## Residual Sum of Squares: 21.414
## R-Squared:    0.67282
## Adj. R-Squared: 0.66276
## F-statistic: 66.8343 on 4 and 130 DF, p-value: < 2.22e-16
```

- Asume que no hay efectos no observados entre países o años.
- Los coeficientes muestran el impacto marginal de cada variable sobre el PIB per cápita.
- No controla por heterogeneidad no observada.

Modelo de Efectos Fijos

```
reg_fe <- plm(lngdp ~ capi + art + labor + Overall.Score, data = df, model = "within")
summary(reg_fe)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = lngdp ~ capi + art + labor + Overall.Score, data = df,
##      model = "within")
##
## Balanced Panel: n = 5, T = 27, N = 135
##
## Residuals:
##      Min.   1st Qu.   Median   3rd Qu.    Max.
## -0.63914 -0.21759 -0.01107  0.21857  0.48446
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## capi           3.5281e-02  1.1074e-02  3.1858  0.00182 **
## art            -1.1075e-04  2.3294e-05 -4.7546 5.342e-06 ***
## labor           1.4928e-07  1.8346e-08  8.1372 3.318e-13 ***
## Overall.Score -2.8506e-02  5.6364e-03 -5.0576 1.462e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    26.608
## Residual Sum of Squares: 11.035
## R-Squared:              0.5853
## Adj. R-Squared:         0.55897
## F-statistic: 44.4579 on 4 and 126 DF, p-value: < 2.22e-16
```

- Controla por factores no observables que son constantes en el tiempo dentro de cada país.
- Ideal si creemos que existen características no medidas (como instituciones o cultura) que afectan al PIB.
- Se enfoca en las variaciones dentro de cada país.

Modelo de Efectos Aleatorios

```
reg_re <- plm(lngdp ~ capi + art + labor + Overall.Score,
              data = df,
              model = "random",
              random.method = "amemiya")
summary(reg_re)
```

```
## Oneway (individual) effect Random Effect Model
##      (Amemiya's transformation)
```

```
##
## Call:
## plm(formula = lngdp ~ capi + art + labor + Overall.Score, data = df,
##      model = "random", random.method = "amemiya")
##
## Balanced Panel: n = 5, T = 27, N = 135
##
## Effects:
##              var std.dev share
## idiosyncratic 0.08488 0.29134 0.045
## individual    1.82044 1.34924 0.955
## theta: 0.9585
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -0.6297500 -0.2265640  0.0045183  0.2299097  0.5437491
##
## Coefficients:
##              Estimate Std. Error z-value Pr(>|z|)
## (Intercept)  7.0840e+00  7.8705e-01  9.0007 < 2.2e-16 ***
## capi         3.6311e-02  1.1078e-02  3.2778  0.001046 **
## art         -9.8744e-05  2.2568e-05 -4.3754  1.212e-05 ***
## labor        1.3807e-07  1.7538e-08  7.8726  3.473e-15 ***
## Overall.Score -2.8911e-02  5.6362e-03 -5.1294  2.906e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    26.675
## Residual Sum of Squares: 11.421
## R-Squared:    0.57184
## Adj. R-Squared: 0.55866
## Chisq: 173.622 on 4 DF, p-value: < 2.22e-16
```

- Supone que las diferencias no observadas entre países no están correlacionadas con las variables explicativas.
- Más eficiente que FE si los supuestos se cumplen.

Comparación de modelos

Test F: ¿Pooled vs. Efectos Fijos?

```
pFtest(reg_fe, reg_pool)
```

```
##
## F test for individual effects
##
## data:  lngdp ~ capi + art + labor + Overall.Score
## F = 29.63, df1 = 4, df2 = 126, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

- H_0 : El modelo Pooled es adecuado (sin efectos individuales).
- H_1 : El modelo de Efectos Fijos es preferido.
- Si el valor p es bajo ($p < 0.05$), se rechaza H_0 se prefiere Efectos Fijos.

Test de Hausman: ¿Efectos Aleatorios vs. Efectos Fijos?

```
phptest(reg_re, reg_fe)
```

```
##
## Hausman Test
##
## data: lngdp ~ capi + art + labor + Overall.Score
## chisq = 4.1114, df = 4, p-value = 0.3911
## alternative hypothesis: one model is inconsistent
```

- H_0 : El modelo de Efectos Aleatorios es consistente.
- H_1 : El modelo de Efectos Fijos es preferido.
- Si el valor p es bajo hay correlación entre efectos no observados y las variables usar FE.

Conclusión

- Si ambos tests favorecen el modelo de Efectos Fijos, se concluye que existen efectos específicos de país que deben ser controlados.
- Si el Test de Hausman no rechaza H_0 , Efectos Aleatorios es más eficiente.
- La elección final del modelo depende del resultado de las pruebas y del objetivo del análisis.