

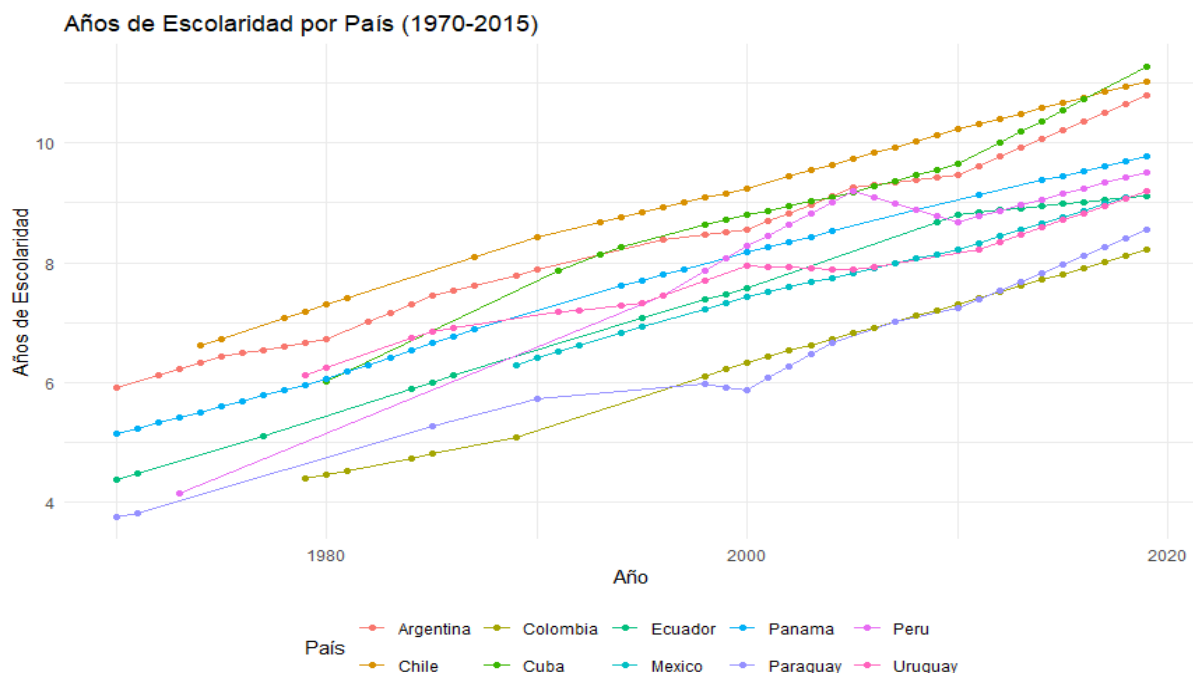
Tercer departamental Panel de datos: actividad

A partir de la base de datos de Barro Lee1 se toman los años de escolaridad promedio de 146 países para el rango de edad de 35 a 44 años para el periodo 1950 a 2015 (variable edu) y se toma del Banco Mundial el gasto público en educación como porcentaje del PIB2 para el mismo periodo de análisis (variable gdp_edu), con la información anterior realice lo siguiente:

1. De la base de datos “panel_educacion.RData” genere un subconjunto de datos tomando los siguientes países de Latinoamérica: Argentina, Brasil, Chile, Colombia, Cuba, Costa Rica, Ecuador, México, Perú, Paraguay, Panamá y Uruguay.

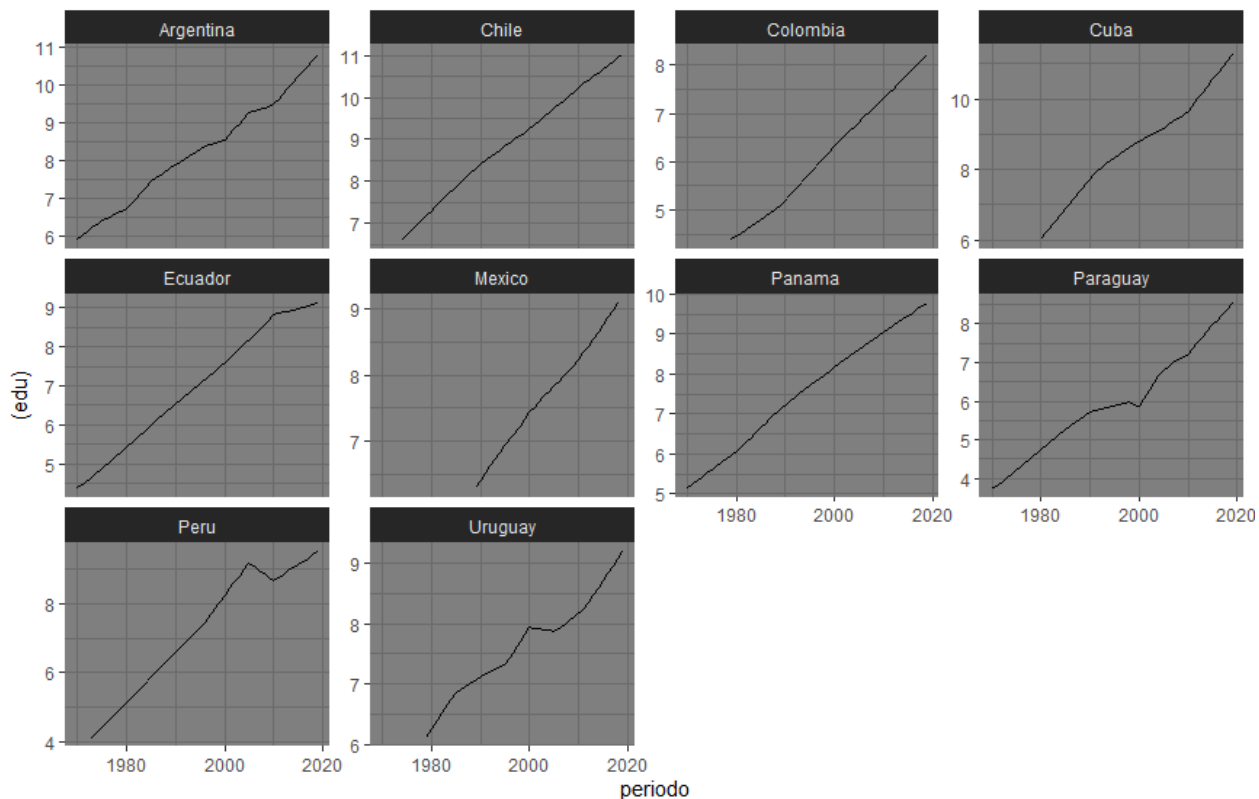
```
> #Filtrar países
> Panel_edu_filtrado <- panel_edu %>%
+   filter(pais %in% c("Mexico", "Argentina", "Brasil", "Chile", "Ecuador", "Peru",
+                     "Paraguay", "Panama", "Uruguay", "Cuba", "Costarica", "Colombia"))
> Panel_edu_filtrado
# A tibble: 283 x 7
  periodo    pais    clave region    ingreso    edu    gdp_edu
  <date>    <chr>    <chr> <chr>    <chr>    <dbl>    <dbl>
1 1970-01-01 Argentina ARG    Latin America & Caribbean Upper middle income 5.92    1.46
2 1972-01-01 Argentina ARG    Latin America & Caribbean Upper middle income 6.12    1.94
3 1973-01-01 Argentina ARG    Latin America & Caribbean Upper middle income 6.22    1.78
4 1974-01-01 Argentina ARG    Latin America & Caribbean Upper middle income 6.33    1.92
5 1975-01-01 Argentina ARG    Latin America & Caribbean Upper middle income 6.43    1.84
6 1976-01-01 Argentina ARG    Latin America & Caribbean Upper middle income 6.49    1.17
7 1977-01-01 Argentina ARG    Latin America & Caribbean Upper middle income 6.54    1.72
8 1978-01-01 Argentina ARG    Latin America & Caribbean Upper middle income 6.60    1.94
9 1979-01-01 Argentina ARG    Latin America & Caribbean Upper middle income 6.66    2.40
10 1980-01-01 Argentina ARG    Latin America & Caribbean Upper middle income 6.72    2.61
# i 273 more rows
# i Use `print(n = ...)` to see more rows
```

2. Grafique los años de escolaridad promedio de todos los países (variable edu) para el periodo de 1970 a 2015 y comente la gráfica.

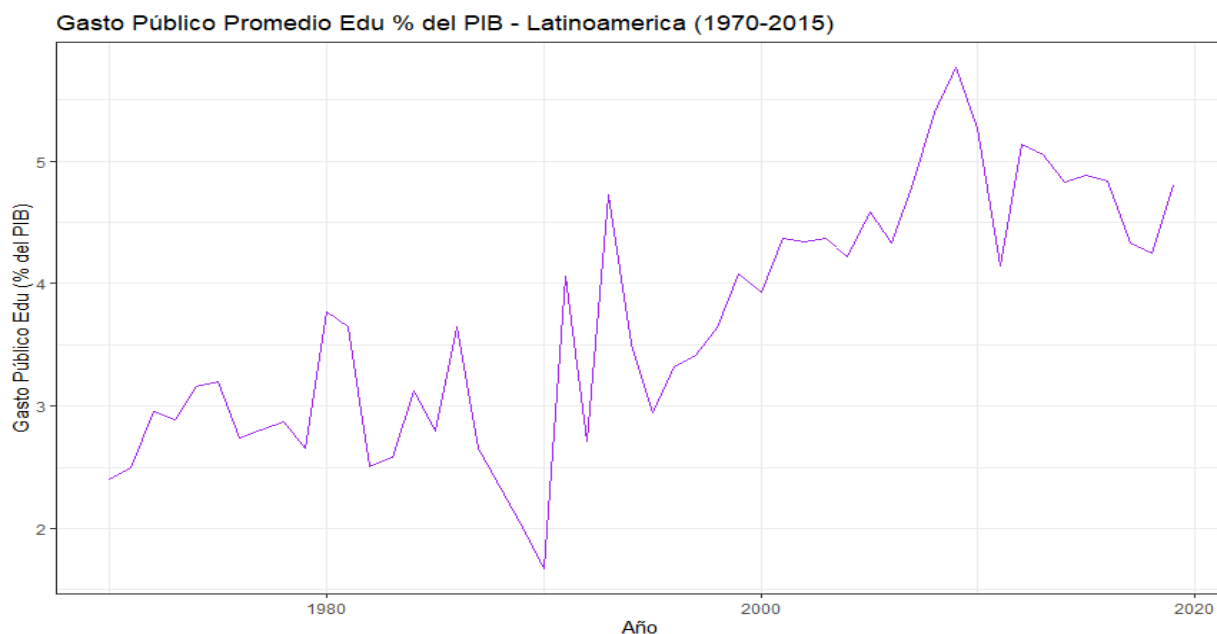


A través de los años la variable educación a tenido un crecimiento constante en la mayoría de los países de nuestra selección

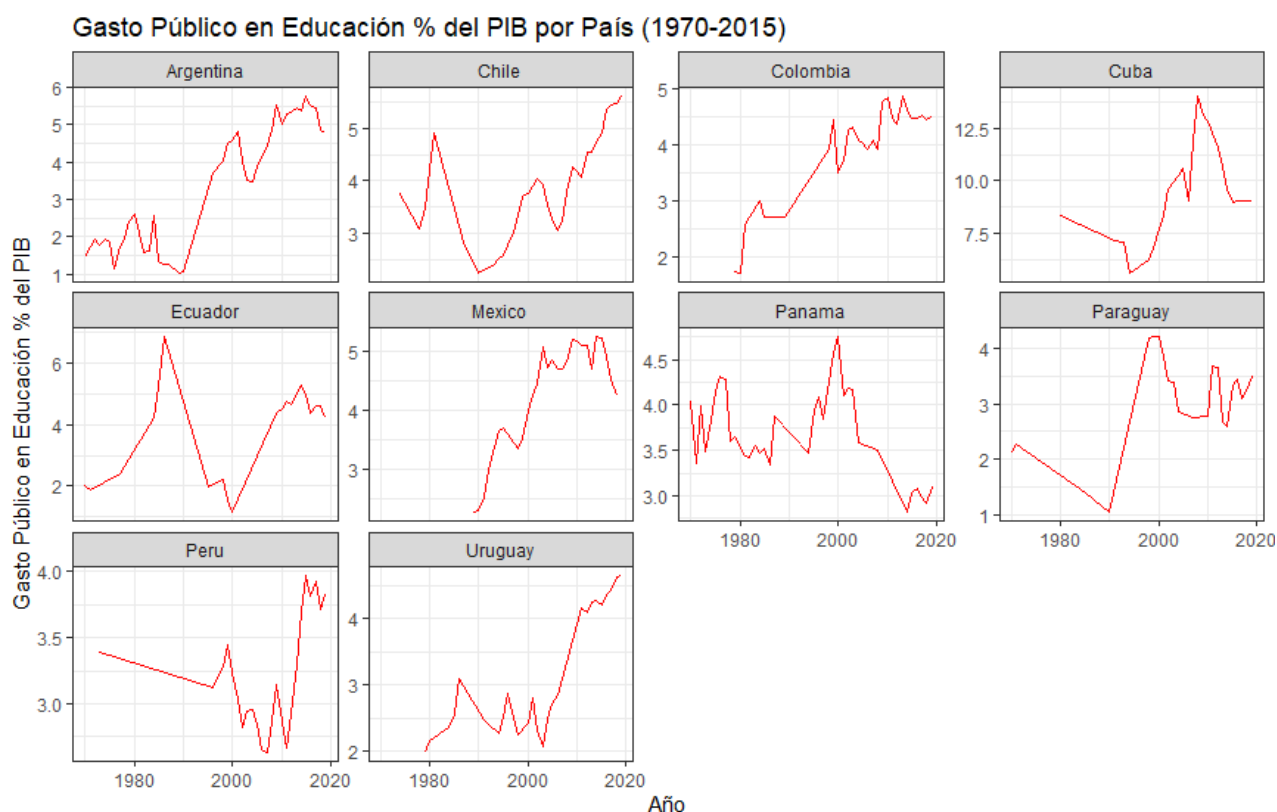
3. Grafique los años de escolaridad promedio para cada uno de los países (variable edu) para el periodo de 1970 a 2015 y comente la gráfica.



4. Grafique el gasto público promedio en educación como porcentaje del PIB de todos los países (variable gdp_edu) para el periodo de 1970 a 2015 y comente la gráfica.



5. Grafique el gasto público promedio en educación como porcentaje del PIB para cada uno de los países (variable edu) para el periodo de 1970 a 2015 y comente la gráfica.



6. Estime el siguiente modelo de panel de datos:

$$eduit = \alpha_i + \alpha igdp_eduit + \epsilon it$$

a) Estime a través de pooling vs efectos fijos. ¿Qué estimación es más eficiente y por qué?

```
> # a) Estimación con Pooling vs Efectos Fijos
> # Pooling
> modelo_pooling <- plm(edu ~ gdp_edu,
+                         data = Panel_edu_filtrado,
+                         model = "pooling",
+                         index = c("pais", "periodo"))
> summary(modelo_pooling)
Pooling Model

Call:
plm(formula = edu ~ gdp_edu, data = Panel_edu_filtrado, model = "pooling",
     index = c("pais", "periodo"))

Unbalanced Panel: n = 10, T = 21-41, N = 283

Residuals:
    Min.    1st Qu.    Median    3rd Qu.    Max.
-3.601304 -0.921907  0.033837  1.079912  2.609092

Coefficients:
              Estimate Std. Error t-value Pr(>|t|)
(Intercept)  6.628092   0.184606  35.904 < 2.2e-16 ***
gdp_edu      0.327023   0.040644   8.046 2.42e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    654.28
Residual Sum of Squares: 531.77
R-Squared:              0.18725
Adj. R-Squared:         0.18435
F-statistic: 64.7379 on 1 and 281 DF, p-value: 2.4202e-14
```

b) Estime a través de efectos fijos vs efectos aleatorios. ¿Qué estimación es más eficiente y por qué?

```

> # Efectos fijos (within)
> modelo_fijos <- plm(edu ~ gdp_edu,
+                     data = Panel_edu_filtrado,
+                     model = "within",
+                     index = c("pais", "periodo"))
> summary(modelo_fijos)
Oneway (individual) effect Within Model

Call:
plm(formula = edu ~ gdp_edu, data = Panel_edu_filtrado, model = "within",
     index = c("pais", "periodo"))

Unbalanced Panel: n = 10, T = 21-41, N = 283

Residuals:
    Min.   1st Qu.   Median   3rd Qu.    Max.
-4.57317 -0.55378  0.21139  0.68298  2.77863

Coefficients:
              Estimate Std. Error t-value Pr(>|t|)
gdp_edu  0.535461    0.056226   9.5234 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    441.6
Residual Sum of Squares: 331.17
R-Squared:              0.25006
Adj. R-Squared: 0.22249
F-statistic: 90.696 on 1 and 272 DF, p-value: < 2.22e-16

> # Prueba F de efectos fijos vs pooling
> pFtest(modelo_fijos, modelo_pooling)

F test for individual effects

data:  edu ~ gdp_edu
F = 18.306, df1 = 9, df2 = 272, p-value < 2.2e-16
alternative hypothesis: significant effects

```

```

> # b) Efectos Fijos vs Efectos Aleatorios
> # Efectos aleatorios
> modelo_aleatorios <- plm(edu ~ gdp_edu,
+                           data = Panel_edu_filtrado,
+                           model = "random",
+                           index = c("pais", "periodo"))
> summary(modelo_aleatorios)
Oneway (individual) effect Random Effect Model
(Swamy-Arora's transformation)

Call:
plm(formula = edu ~ gdp_edu, data = Panel_edu_filtrado, model = "random",
     index = c("pais", "periodo"))

Unbalanced Panel: n = 10, T = 21-41, N = 283

Effects:
              var std.dev share
idiosyncratic 1.2175  1.1034 0.628
individual    0.7215  0.8494 0.372
theta:
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.7273  0.7437  0.7575  0.7643  0.7884  0.8012

Residuals:
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-4.2690 -0.6092  0.2297  0.0091  0.7342  2.6762

Coefficients:
              Estimate Std. Error z-value Pr(>|z|)
(Intercept)  5.844024   0.355175  16.4539 < 2.2e-16 ***
gdp_edu      0.500261   0.053049   9.4302 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    460.72
Residual Sum of Squares: 344.78
R-Squared:              0.25209
Adj. R-Squared:         0.24943
Chisq: 88.9282 on 1 DF, p-value: < 2.22e-16

```

```

> # Prueba de Hausman
> phtest(modelo_fijos, modelo_aleatorios)

      Hausman Test

data:  edu ~ gdp_edu
chisq = 3.5694, df = 1, p-value = 0.05885
alternative hypothesis: one model is inconsistent

```

c) Realice las pruebas de auto correlación y heterocedasticidad y mencione si el modelo pasa dichas pruebas.

```
> # Prueba de heterocedasticidad (Breusch-Pagan)
> bptest(modelo_fijos)

studentized Breusch-Pagan test

data: modelo_fijos
BP = 0.32556, df = 1, p-value = 0.5683

> # Prueba de autocorrelación (Breusch-Godfrey)
> bgtest(modelo_fijos)

Breusch-Godfrey test for serial correlation of order up to 1

data: modelo_fijos
LM test = 186.66, df = 1, p-value < 2.2e-16
```

d) Si el modelo presenta problemas de estimación corrija el modelo.

```
> # Varianzas robustas tipo Arellano (HAC para panel)
> vcov_robust <- vcovHC(modelo_fijos, method = "arellano", type = "HC1", cluster = "group")
> # Reestimación con errores estándar robustos
> coeftest(modelo_fijos, vcov. = vcov_robust)

t test of coefficients:

      Estimate Std. Error t value Pr(>|t|)
gdp_edu  0.53546    0.13153   4.071 6.143e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

e) Una vez determinado el mejor método de estimación y sus correcciones, interprete los resultados.

```
> # Medias generales (toda la base filtrada)
> Panel_edu_filtrado %>%
+   summarise(media_edu_total = mean(edu, na.rm = TRUE),
+             media_gdp_edu_total = mean(gdp_edu, na.rm = TRUE))
# A tibble: 1 x 2
  media_edu_total media_gdp_edu_total
      <dbl>          <dbl>
1          7.96            4.07
```

Entregar sus resultados en Word con anexos, el script de R y su correspondiente base de datos. Valor 3 puntos.