Data Selection

When selecting data, we were careful to choose data that is well associated with the terms of the given formula. We chose metrics that avoid bias and are up to date. Here's a list of the metrics we chose and our reasoning behind it:

- GDP per capita (constant 2015 USD)

  With GDP per capita we had a lot of different options, but to eliminate any bias related to inflation we chose the dataset that is constantly expressed in 2015 USD.

- Gross saving (% of total GDP)

  The Solow model requires the saving rate to be expressed as as a percentage of total expenditure, or in other words, percentage of total GDP, so this metric is well suited.

- Population growth (annual %)

  For the last metric, we needed something that captures the per year change in population, hence the annual population growth in percentages was best suited.

Data Processing

Here are the steps we took for preprocessing our data to make it suitable for OLS:

1. Converted individual datasets from wide to long to wide format using the melt function.
2. Merged the three data tables into a single table, keeping country and year consistent.
3. Since there was a lot of missing data between the three datasets, we needed to efficiently drop some of the data that was unusable. We decided to drop any country/year that had more than 30% of its data missing because this way we could balance data retention and quality. With a higher threshold, say 50%, too much unavailable data remained in the data set and imputing the values would introduce some bias. A lower, threshold, like 10%, would discard too much useful data and would make the model less accurate.
4. For the few remaining unavailable data, we imputed the data using the mean value for each country respectively.
5. Since our model has a log-log functional form, we removed any rows with negative values.

We ended up with 81 countries, 48 years, and 3763 total observations in the final dataset, enough to run a satisfactory OLS.

Interpretation of Regression Results

$$\ln gdppc_t = \alpha + \beta \ln s_t - \gamma \ln n_t$$

```
Call:
lm(formula = log(GDP_per_capita) ~ log(savings_rate) + log(population),
    data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-4.1044 -0.8027  0.0775  0.7536  3.7504

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        6.03808    0.12877   46.89   <2e-16 ***
log(savings_rate)  0.96304    0.04224   22.80   <2e-16 ***
log(population)   -0.79034    0.02016  -39.21   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.148 on 3760 degrees of freedom
Multiple R-squared:  0.3765,    Adjusted R-squared:  0.3762
F-statistic:  1135 on 2 and 3760 DF,  p-value: < 2.2e-16
```

- The t-tests suggest all coefficients in the model are significant at the 1% level.
- The coefficients are jointly significant at the 1% level, as shown by the F-statistic.
- Our empirical data suggests the following values for the model coefficients
  - **α = 6.038**
  - **β = 0.963**
  - **γ = 0.790**
- Inline with the model, our coefficients suggest the following relationship:
  - A **1% increase** in the savings rate is associated with a **0.963% increase** in GDP per capita.
  - A **1% increase** in population growth is associated with a **0.79% decrease** in GDP per capita
  - **Positive** intercept term, representing other growth factors.
- R^2 is rather low, which shows that the Solow-Swan model is quite limited and suggests that other factors also influence GDP.

Adding education as an additional regressor

$$\ln gdppc_t = \alpha + \beta \ln s_t - \gamma \ln n_t + \delta \ln e_t$$

```
Call:
lm(formula = log(GDP_per_capita) ~ log(savings_rate) + log(population) +
    log(education), data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-2.09277 -0.40299  0.00664  0.36743  1.41369

Coefficients:
                  Estimate Std. Error t value
(Intercept)        3.83105    0.30624  12.510
log(savings_rate)  0.50003    0.09926   5.038
log(population)   -0.28204    0.03213  -8.778
log(education)     2.86195    0.08957  31.951
                  Pr(>|t|)
(Intercept)        < 2e-16 ***
log(savings_rate) 6.33e-07 ***
log(population)    < 2e-16 ***
log(education)     < 2e-16 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.608 on 572 degrees of freedom
Multiple R-squared:  0.7421,	Adjusted R-squared:  0.7408
F-statistic: 548.7 on 3 and 572 DF,  p-value: < 2.2e-16
```

- We decided to add **government spending on education as % of GDP** as an additional regressor.
- The t-tests suggest all coefficients in the new model are significant at the 1% level.
- The coefficients are jointly significant at the 1% level, as shown by the F-statistic.
- Our empirical data suggests the following values for the model coefficients

  α = 3.831

  β = 0.500

  γ = 0.282

  δ = 2.862
- Inline with the model, our coefficients suggest the following relationship:
  - A **1% increase** in the savings rate is associated with a **0.500% increase** in GDP per capita.
  - A **1% increase** in population growth is associated with a **0.282% decrease** in GDP per capita
  - A **1% increase** in education spending is associated with a **2.862 % increase** in GDP per capita
  - **Positive** intercept term, representing other growth factors.

- R^2 is a lot higher now, which shows that the new model gives a better explanation of the factors that influence GDP.
- The residual standard error decreased from 1.148 to 0.608. This means that the model’s predictions are now more precise.

Interpretation of the new results

Adding education as an additional regressor adjusts the estimated coefficients of savings and population because of **Omitted Variable Bias**. In our old model without education, savings and population are absorbing some of the effects that should be attributed to education. This is due to the fact that education is correlated with both savings and GDP per capita. The old model without education wrongfully attributed some of the effects of education to savings and population. In the new model with education included these misattributions are corrected and we get better estimates.

Savings & Education:

- Savings and Education are positively correlated. Countries that save more tend to invest more in education.The economical intuition behind this is that countries that prioritize savings rather than initial consumption gain greater economic stability in the medium run. This stability allows for long-term investments in human capital development.
- The old model overestimated the impact of savings because it was partly capturing the effect of education.
- In the new model the savings coefficient drops from 0.96 to 0.50 because now the model separates the true effects of education and savings.

Population & Education:

- Population growth and education are negatively correlated. Higher population growth is often associated with lower investment in education. One reason for this is that rapid population expansion accelerates the obsolescence of infrastructure. Governments must continuously reinvest just to maintain existing standards rather than improving human capital.
- In the old model population growth seemed to reduce GDP per capita significantly.
- After adding education, the coefficient shrinks from -0.79 to -0.28, because some of the negative effect of population growth was actually due to lower education levels.

Education & GDP:

- The new model shows the importance of investing in education.
- A 1% increase in education spending is associated with an 2.86% increase in GDP per capita.
- This finding supports the economic theory that human capital development drives productivity and long-term growth.