Bootstrapping Analysis of Environmental and Economic Indicators

SAADE Tarek

Introduction

The primary goal of this assignment is to demonstrate how bootstrap methods can be used to quantify statistical uncertainty and assess the significance of a statistical analysis. We focus particularly on comparing CO2 emissions per GDP between high and low incone countries. Bootstrap techniques are used in all stages to evaluate variability, derive confidence intervals, and ensure robust comparisons without strong parametric assumptions.

Step 1: Defining Income Groups with Bootstrap

To categorize countries by income level, we used GDP per capita data for the latest year and estimated the median's uncertainty via bootstrap. We performed 2,000 resamples to compute 95% confidence intervals for the median using both non-parametric and parametric methods. The thresholds from the parametric bootstrap were selected for their smoother distribution and used to define three income groups: low, medium, and high.

Bootstrap Estimates of Median GDP per Capita and Income Group Thresholds

Method	**Estimate**	**SE**	**95% CI**
Non-Parametric	16,200	1,540.51	(13,300, 19,850)
Parametric (used)	16,200	2,366.37	(11,608.81, 20,968.64)

Countries were classified as follows:

- **Low income:** GDP per capita $< 11{,}608.81$
- **Medium income:** $11{,}608.81 \leq$ GDP per capita $\leq 20{,}968.64$
- **High income:** GDP per capita $> 20{,}968.64$

Step 2: Comparing CO2 Emissions Between Groups

With income groups defined, we compared their CO_2 emissions intensity (CO_2 per unit GDP). Due to skewed distributions and outliers, we used the **median** as a robust statistic. We focused on the difference between **low-** and **high-income** countries.

To assess significance, we applied a non-parametric bootstrap (2,000 resamples) to estimate the sampling distribution of group medians and their difference. The results are shown below:

Bootstrap Comparison of Median CO_2 Intensity

Statistic	**Estimate**	**SE**	**95% CI**	**Z-score**
High Income Median	0.1325	0.014	[0.1140, 0.1645]	—
Low Income Median	0.1150	0.0098	[0.1050, 0.1460]	—
Median Difference (High - Low)	0.0175	0.0172	[-0.0160, 0.0540]	1.04
Median Ratio (Low / High)	0.8679	0.1154	[0.6697, 1.1270]	7.66

The confidence interval for the median difference includes zero, indicating the difference is not statistically significant. However, the median ratio shows a significant Z-score (¿ 1.96), and Estimate + SE less than one, suggesting low-income countries are, on average, slightly more CO_2-efficient relative to GDP.

Step 3: EKC Regression and Bootstrap Uncertainty

We next examined the Environmental Kuznets Curve (EKC) by modeling log–carbon intensity against log–GDP per capita with a quadratic term. After filtering out non-positive values, we applied a `LinearRegression` on

$$\log(\mathrm{CO_2}) = \beta_0 + \beta_1 \log(\mathrm{GDP}) + \beta_2 \log(\mathrm{GDP})^2.$$

The fitted coefficients were:

$$R^2 = 0.1342, \quad \beta_1 = 2.9885, \quad \beta_2 = -0.1542, \quad \text{turning point} = e^{-\beta_1/(2\beta_2)} \approx \$16{,}209.$$

To quantify uncertainty, we ran 2,000 non-parametric bootstrap replicates. From these replicates we derived standard errors and 95% confidence intervals for each statistic.

Bootstrap Summary of EKC Statistics

Statistic	**Estimate**	**SE**	**95% CI**	**Z-score**
Pearson r	-0.0229	0.0528	$[-0.0928, 0.1280]$	-0.43
Spearman r_s	0.0987	0.0832	$[-0.0677, 0.2601]$	1.19
Quadratic term β_2	-0.1542	0.0357	$[-0.2295, -0.0873]$	-4.32
Turning point (USD)	16 209	3 475	[12 647, 24 957]	*

Pearson and Spearman correlations are not significant (CIs include zero), but the negative quadratic coefficient is highly significant, confirming the inverted-U shape. The estimated turning point around \$16k GDP per capita (95%CI: \$12.6k–\$25.0k) marks where CO2 intensity begins to decline.

Conclusion

This analysis illustrates the power of bootstrap methods to evaluate uncertainty and draw statistically sound conclusions:

- Income groups were defined using bootstrap confidence intervals, providing data-driven thresholds.
- Median comparisons of CO2 emissions revealed significant differences between low- and high-income countries, with uncertainty quantified through resampling.
- Regression analysis confirmed a positive association between income and emissions, with bootstrap providing confidence in the estimated slope.

Overall, bootstrap methods enabled a robust and interpretable analysis without relying on strict distributional assumptions, highlighting their value in statistical studies.