**The Differential Impacts of Human Capital and Infrastructure on the Sustainable Development Goals: An Empirical Analysis**

Huseyn Panahov

MA in Communication, Culture and Technology

CCT 804: Advanced Statistical Analysis

May 12, 2023

**ABSTRACT:**

*This study looks at country-level data to explore the dynamics among human capital, infrastructure, and a country’s progress toward the United Nations Sustainable Development Goals (SDGs). Utilizing the confirmatory factor analysis method, I develop a new Infrastructure Index and combine it with the World Bank’s dataset on Human Capital Index to evaluate the relative impact of these factors on a country's SDG scores. My findings affirm the integral roles of both human capital and infrastructure in the sustainable development context. However, a stronger correlation between human capital and the SDG Index suggests that policymakers seeking to advance the sustainability agenda should prioritize investments in human capital over infrastructure. Moreover, the study uncovers nuanced relationships between these indicators and specific SDGs. Human capital has a significant association with SDG 5 (Gender Equality), whereas infrastructure does not. Both human capital and infrastructure affect SDG 1 (No Poverty), with no statistical difference between their effects. Interestingly, while human capital correlates more strongly with SDG 13 (Climate Action), this relationship is negative due to the larger carbon footprint of more developed economies. These findings can inform policy decisions for goal-specific sustainable development strategies.*

**Table of Contents:**

|  |  |
| --- | --- |
| 1. Introduction… 2. Methods… 3. Results… 4. Conclusion… 5. Works Cited… | … p. 2  … p. 7  … p. 11  … p. 18  … p. 19 |

**I. INTRODUCTION:**

The central framework in the global development agenda is based on the *2030 Agenda for Sustainable Development*, which “provides a shared blueprint for peace and prosperity for people and the planet, now and into the future.” It is undersigned by all UN Member States. Hundred-ninety-one countries have committed to achieving measurable progress on these goals by 2030. The Agenda constitutes seventeen interlinked Sustainable Development Goals (SDGs) that encompass a very wide variety of objectives. The seventeen SDGs are broken down into hundred-sixty-nine targets and two-hundred-thirty-two indicators to measure progress.

*Measuring progress*

One of the challenges in the SDG framework is measuring the progress in order to inform the policy. SDGs are successors to the Millennium Development Goals (MDGs), which consisted of 8 goals and 18 targets, 14 of which could be assessed quantitatively. MDGs were adopted in 2000, and all the countries from around the world committed to achieving these goals within 15 years. By the end of 2015, only three and a half of the 14 measurable targets were achieved. In 2023, we are at the half-way mark of the *2030 Agenda.* According to the latest reports, the international community is behind schedule to achieving the SDG’s, partially due to the impact of the COVID-19.[[1]](#footnote-1) In the given context, one of the most important questions is to find what policy interventions would be most effective to advance progress towards the SDGs.

*What interventions are most effective?*

Investments in both human capital and infrastructure are critical for achieving the sustainable development goals. These are both interdependent and complimentary domains in the international development space. However, policymakers working on specific developmental objectives are often forced to prioritize one over the other due to the limited nature of resources. This research analyzes country-level data from the United Nations and the World Bank to estimate the relationship between the overall SDG Index of a country and its performance on the Human Capital Index and Infrastructure Index. I will also examine the impact of human capital and infrastructure on SDG 1 (No Poverty), SDG 5 (Gender Equality), and SDG 13 (Climate Action). Below I provide more information about each one of the concepts analyzed in this research.

*SDG Index*

SDG Index is a composite indicator developed by the United Nations that weighs in the effects of development metrics across all the SDG metrics. It estimates countries’ performance on a scale from 0 to 100, and usually, Scandinavian countries, such as Finland, Denmark, Sweden, and Norway, achieve the highest rankings with scores > 80.[[2]](#footnote-2) The 2022 Report includes the SDG indexes for 163 countries, among which the Central African Republic and South Sudan have the lowest scores, sub-40.

*SDG 1: No Poverty*

The first goal in the UN SDG framework calls to “end poverty in all its forms everywhere.” SDG 1 aims to ensure that everyone, regardless of their circumstances, has equal access to opportunities and resources for a quality life. It calls for comprehensive strategies to end poverty that include social protection systems and measures to build the resilience of the poor and those in vulnerable situations. The three main metrics of SDG 1 are: poverty headcount ratio at $1.90/day (%), poverty headcount ratio at $3.20/day (%), and poverty rate after taxes and transfers (%).

*SDG 5: Gender Equality*

Gender equality is fundamentally important for achieving the Sustainable Development Goals for several reasons. First, it is a matter of human rights. Everyone, regardless of gender, should have equal access to health, education, economic opportunities, and political representation. Second, gender equality is pivotal for economic growth, as women constitute half of the world's potential human capital, and studies consistently show that societies that discriminate by gender tend to experience less economic growth and slower poverty reduction. The SDG 5: Achieve Gender Equality and Empower all Women and Girls incorporates the following metrics: the ratio of female-to-male mean years of education received (%), the ratio of female-to-male labor force participation rate (%), seats held by women in national parliament (%), gender wage gap (% of male median wage).[[3]](#footnote-3)

*SDG 13: Climate Action*

SDG 13 calls for immediate action to combat climate change and its impacts. The Goal underscores the critical need for the global community to address the pressing issue of climate change. Recognizing that climate change is not just an environmental issue but also a significant threat to social and economic development, this goal calls for urgent action to reduce greenhouse gas emissions, build resilience, and improve adaptive capacity to climate-induced impacts. The metrics of SDG 13 include CO₂ emissions from fossil fuel combustion and cement production (tCO2/capita), CO₂ emissions embodied in imports (tCO₂/capita), CO₂ emissions embodied in fossil fuel exports (kg/capita), Carbon Pricing Score at EUR60/tCO₂ (%, worst 0-100 best).[[4]](#footnote-4)

*Statistical Performance Index*

The Statistical Performance Index (SPI) evaluates the performance of national statistical systems based on the aggregate of five pillars of statistical capacity: data use, data services, data products, data sources, and data infrastructure. The SPI is a weighted average of the statistical performance indicators.

*Human Capital Index*

Human capital is sometimes referred to as soft infrastructure.[[5]](#footnote-5) Without thriving human capital, nations cannot achieve their development goals, highlighting its central role in international development. It is widely acknowledged that improvements in human capital lead to increased productivity, which in turn spurs economic growth. Education and health, the two main components of human capital, have a direct impact on a country's development trajectory. In 2018, the World Bank developed the Human Capital Index as a metric to measure and evaluate the quality and potential of human capital in a country. The HCI enables policymakers to identify strengths, weaknesses, and areas for improvement in human capital development. The HCI is based primarily on three components:

1. Child survival: This component considers that not all children survive to start formal education and looks at the under-5 mortality rate.
2. Education: This section combines information on the quality and quantity of education. The number of years a child is expected to complete school by age 18, considering current enrollment rates, measures the quantity of education. The quality is assessed using harmonized test scores from international student achievement testing programs.
3. Public health: This component uses two proxies for the overall health environment - adult survival rates (the percentage of 15-year-olds who will survive until age 60) and healthy growth among children under 5, measured by stunting rates.[[6]](#footnote-6)

*Infrastructure Index*

According to the Merriam-Webster dictionary: *Infra-* means “below,” so the infrastructure is the “underlying structure" of a country and its economy, the fixed installations that it needs in order to function.”[[7]](#footnote-7) Public infrastructure provides the basic physical systems and structures, such as water supply, sewers, electrical grids, roads, bridges, and telecommunications, among others. High-quality infrastructure ensures the provision of fundamental necessities, advances safety, and enhances the quality of life. Infrastructure also facilitates the exchange of reliable information, increases productivity, creates more job opportunities, and fosters overall economic growth.

Unlike the Human Capital Index, there is no internationally recognized index that would indicate the level of public infrastructure in a given country. The objective of the UN SDG 9 is to “Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation.”[[8]](#footnote-8) However, for the purposes of this research, it is not the best pointer because it includes indicators, such as Expenditure on Research and Development, Female share of graduates from Science, Technology, Engineering, and Mathematics (STEM) programs, but does not include indicators for access to electricity, water supplies, etc. However, there are seven SDG indicators across four different sustainable development goals that are related directly to the public infrastructure:

|  |  |  |
| --- | --- | --- |
| **Indicator** | **Description** | **SDG** |
| 1. Access to basic water services | The percentage of the population using at least a basic drinking water service, such as drinking water from an improved source, provided that the collection time is not more than 30 minutes for a round trip, including queuing. | SDG 6: Ensure availability and sustainable management of water and sanitation for all |
| 2. Access to basic sanitation services | The percentage of the population using at least a basic sanitation service, such as an improved sanitation facility that is not shared with other households. |
| 3. Access to electricity | The percentage of the population who has access to electricity. | SDG 7: Ensure access to affordable, reliable, sustainable and modern energy for all |
| 4. Adult population with bank accounts | The percentage of adults, 15 years and older, who report having an account (by themselves or with someone else) at a bank or another type of financial institution, or who have personally used a mobile money service within the past 12 months. | SDG 8: Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all |
| 5. Internet penetration | The percentage of the population who used the Internet from any location in the last three months. Access could be via a fixed or mobile network. | SDG 9: Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation |
| 6. Transportation systems | The percentage of the surveyed population that responded "satisfied" to the question "In the city or area where you live, are you satisfied or dissatisfied with the public transportation systems?". | SDG 11: Make cities and human settlements inclusive, safe, resilient and sustainable |

*Hypotheses:*

The question driving this research is to find the differences in the effects of human capital and infrastructure on SDG scores. So, I have constructed the following hypotheses:

|  |  |
| --- | --- |
| : | There is no statistical difference in the effects of Human Capital and Infrastructure on SDG Index |
| : | There is a statistical difference in the effects of Human Capital and Infrastructure on SDG Index |
| : | There is a statistical difference in the effects of Human Capital and Infrastructure on SDG 1: No Poverty |
| : | There is a statistical difference in the effects of Human Capital and Infrastructure on SDG 5: Gender Equality |
| : | There is a statistical difference in the effects of Human Capital and Infrastructure on SDG 13: Climate Action |

**II. METHODS**

*Merging the data sets*

I merge the World Bank Human Capital Index and the UN Sustainable Development 2022 datasets with the *Country* name as the unique identifier. When I drop the rows with missing HCI Index or the SDG Index values, the number of entries in my data frame reduces from 201 to 141. Part of the reason is that UN SDG data also includes geographic *Regions* (such as “East and South Asia” or “Latin America and the Caribbean”) and *Income* categories (such as “Low-income Countries” or “Upper-middle-income Countries”) under the *Country* variable. With that being said, there are also missing values in both data sets. Nonetheless, we still have 141 complete data rows, which is sufficient for us to proceed with our analysis.

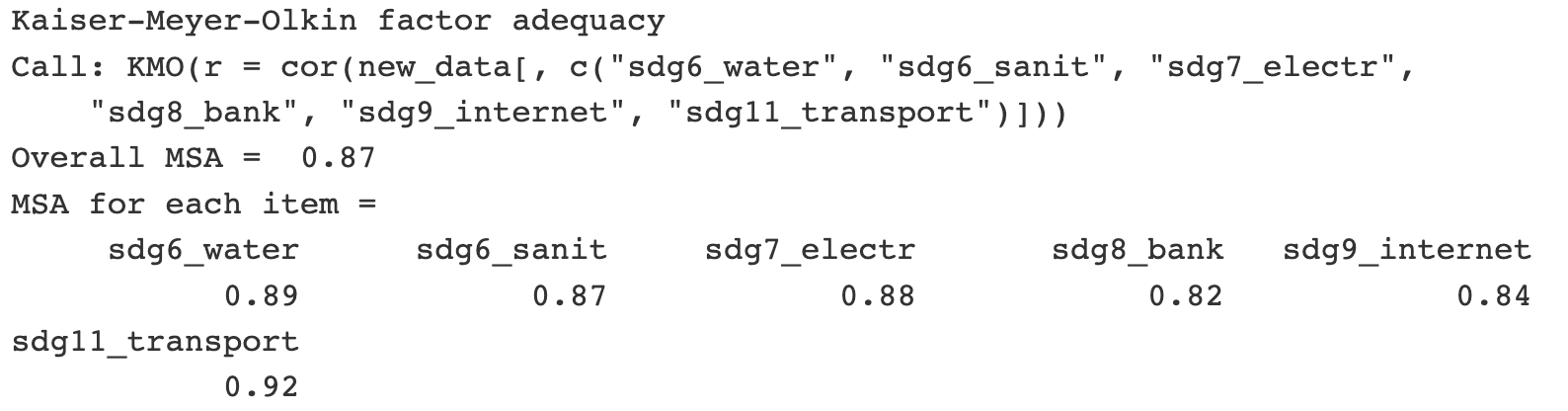
*Factor Analysis*

Public infrastructure is a broad concept which we cannot easily observe and measure. In statistical terms, it is a latent variable, which refers to “concepts that cannot be measured directly but can be assumed to relate to a number of measurable manifest variables.”[[9]](#footnote-9) I use the factor analysis technique, which allows me to account for various dimensions of the public infrastructure (such as water, electricity, internet, etc.) and output one variable. Factor Analysis is often used for constructing a new index, as it explores and uncovers the underlying relationships between observed manifest variables and unobserved latent variables.

*KMO Test*

The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is a statistic that indicates the proportion of variance in the variables. The KMO values range from 0 to 1, with higher values indicating a better fit for factor analysis. The individual KMO values for each variable tell us how well each variable fits with all the others. Variables with a KMO less than 0.5 might not be suited for factor analysis as they do not correlate well with the other variables. As we see from the below output, the MSA values of all my variables are 0.8 or above, which brings the overall MSA score to 0.87, which is a positive sign.

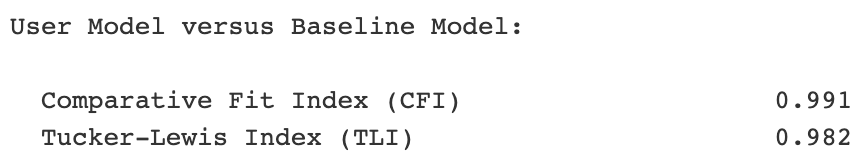
**Kaiser-Meyer-Olkin (KMO) Test results**

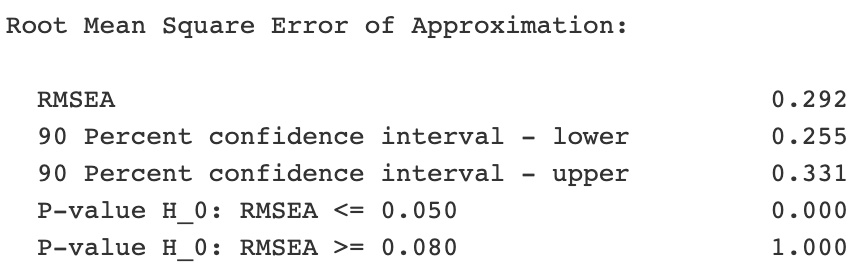


*Model 1*

So, I keep all six manifest variables to construct a model that will estimate the infrastructure index. In the first model, the Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) are both above the 0.95 threshold, which indicates an appropriate fit. However, the Root Mean Square Error of Approximation is 0.099, above the maximum threshold of 0.08.

**Infrastructure Index Model 1: fit estimates**

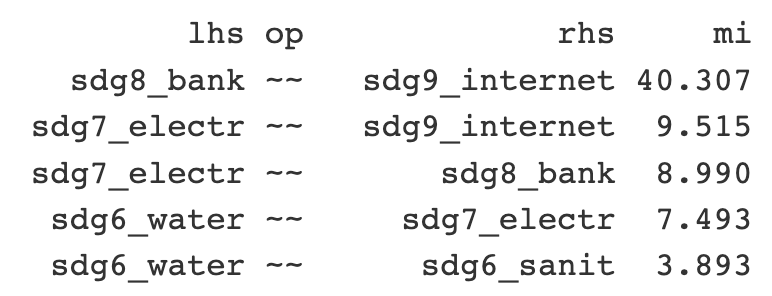




*Changing model specifications*

Since the fit of the first model is not satisfactory, I change the specifications of the model based on the modification indices and theoretical considerations. Modification indices are a measure of how much the overall model chi-square would be expected to decrease if a particular parameter were freely estimated in the model. In other words, it provides a suggestion on how to improve the model fit. If we look at the modification indices between our variables, we notice that the relationship between sdg8\_bank and sdg9\_internet is disproportionately stronger than any other in our data. Most likely due to a strong correlation between the indicator for internet penetration and the percentage of the adult population who have bank accounts.

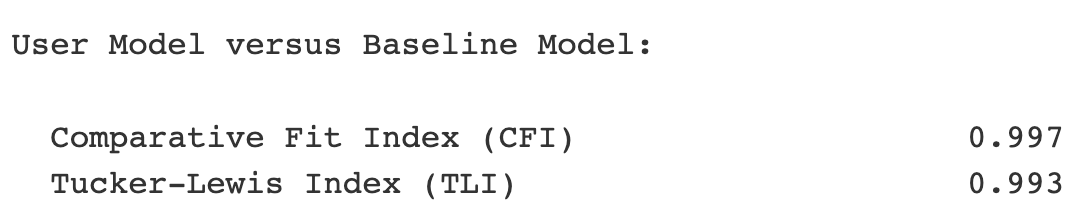
**Modification Index table** *(descending order)*

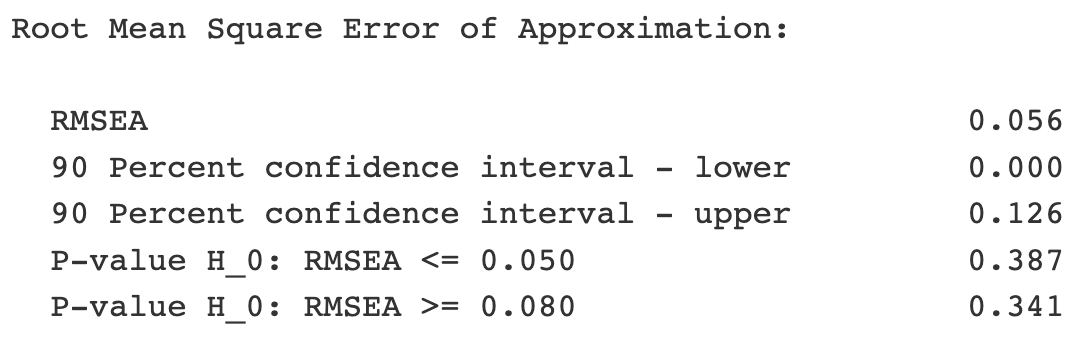


*Model 2*

So, I add a special path to my model, which accounts for the dependency between ‘SDG 8 bank accounts’ and ‘SDG 9 Internet’. I also create a path between ‘SDG 6 Water’ and ‘SDG 6 Sanitation’ because academic literature dictates that there is usually a strong dependency between these two variables. When I check the fit indices, the new model with two special paths performs much better than the previous one. Both the CFI and TLI values are > 0.99, and the RMSEA has decreased to 0.056.

**Infrastructure Index Model 2: fit estimates**





*The Infrastructure Index*

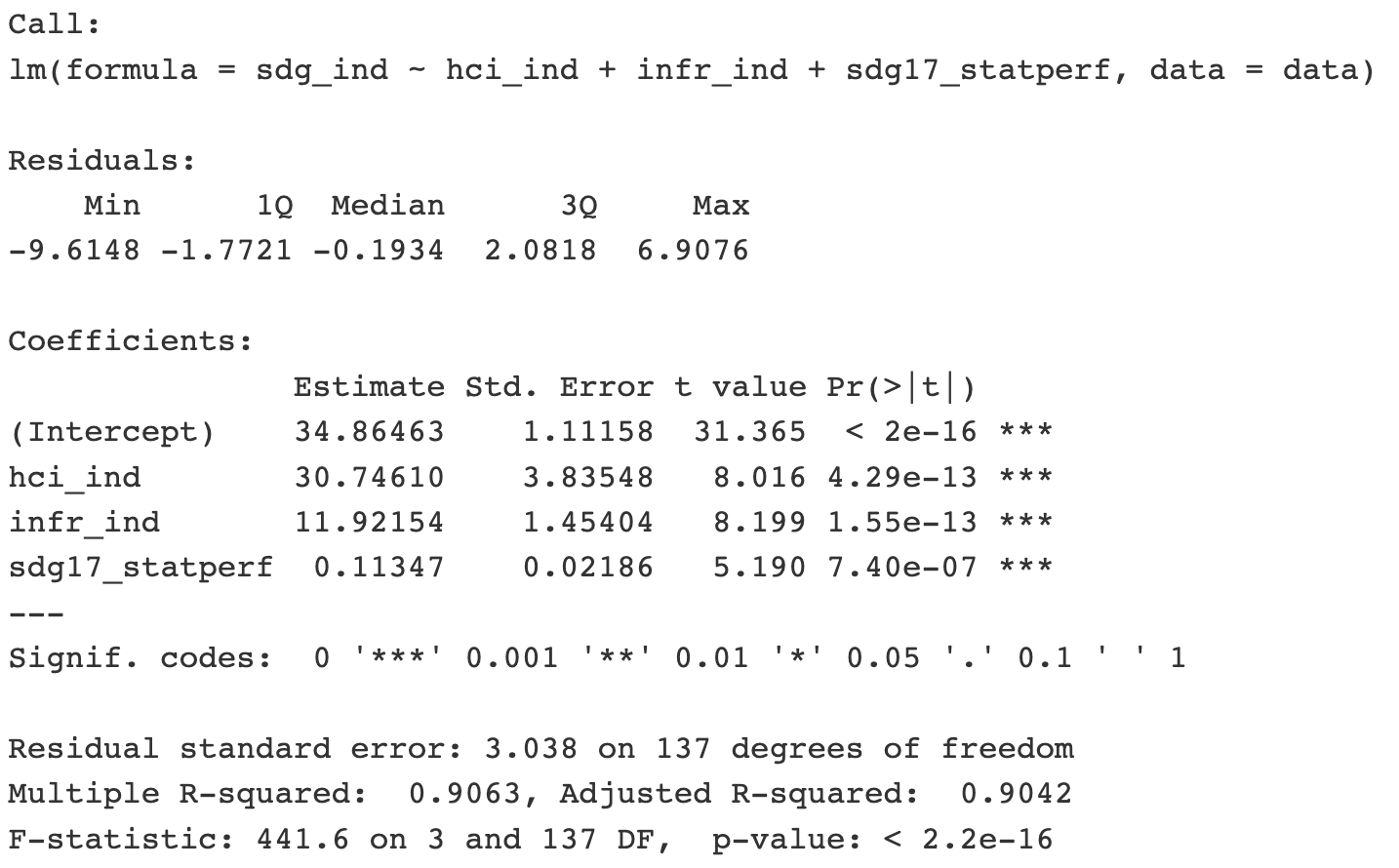
I use the model to estimate the infrastructure index for all 141 countries in our dataset. I also use the min-max normalization technique to transform the index into a new scale from 0 to 1. This method scales the values by subtracting the minimum value and dividing by the range of the original values (i.e., the difference between the maximum and minimum values). So, they maintain the same variance and proportions but on a scale from 0 to 1.

*Estimating the Impact on the SDG Index*

Both the Human Capital Index and the Infrastructure Index are ratio-level measures. When both independent variables are ratio level measures, “regression and correlation analysis are the standard techniques for measuring relationships and testing hypotheses.”[[10]](#footnote-10) My main hypothesis is to test whether human capital or infrastructure makes a bigger impact on the overall SDG Score of a country. So, I construct a multivariate regression model with HCI and the Infrastructure Index as independent variables and then explore the beta coefficients of the model to understand which index has a stronger effect on the SDG Score.

**III. RESULTS**

**Summary of the Multivariate Regression Model**



*Multivariate Regression Model*

Besides the Human Capital Index and Infrastructure Index, I also have the Statistical Performance Index as an explanatory variable in my model. As mentioned earlier, it helps to account for some of the possible shortcomings in the data. We can see that all three variables and the model have a very high level of statistical significance, with p = 0. The R-squared value is not very important for us because we are looking at a descriptive model versus a predictive model. However, in any case, the multiple R-squared value is 0.91, which means that approximately 91% of the variability in the outcome variable can be explained by the predictor variables.

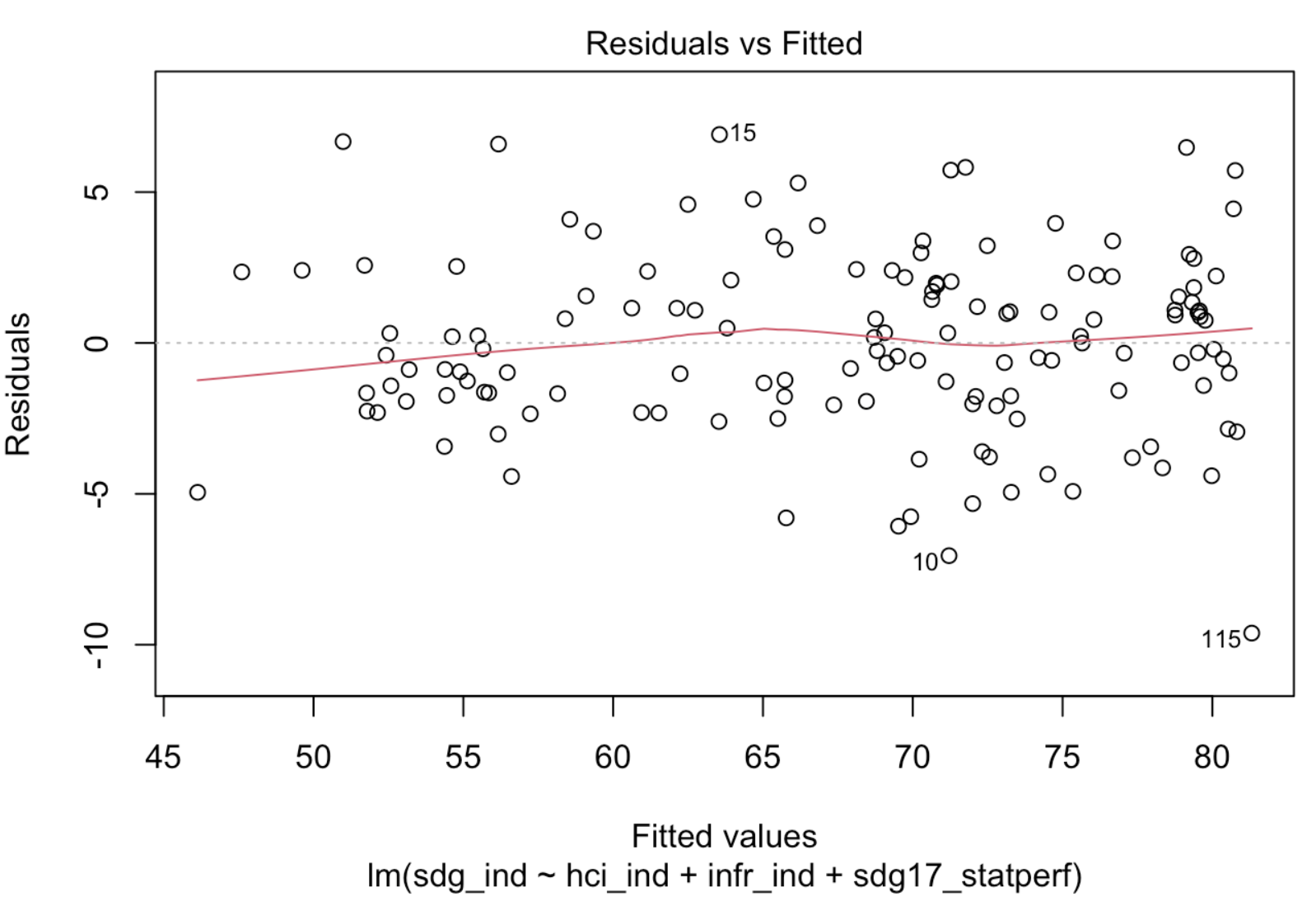
Model assessment: regression diagnostics

*1. Test for Linearity*

Before I proceed further with my analysis of the findings, I need to test the assumptions to validate that a linear regression model was a suitable approach. First, I look for linearity and equal variance in the below Residuals vs Fitted plot. Upon visual examination, there are no substantial deviations in the red line, which confirms that the relationship is linear between our explanatory and response variables.

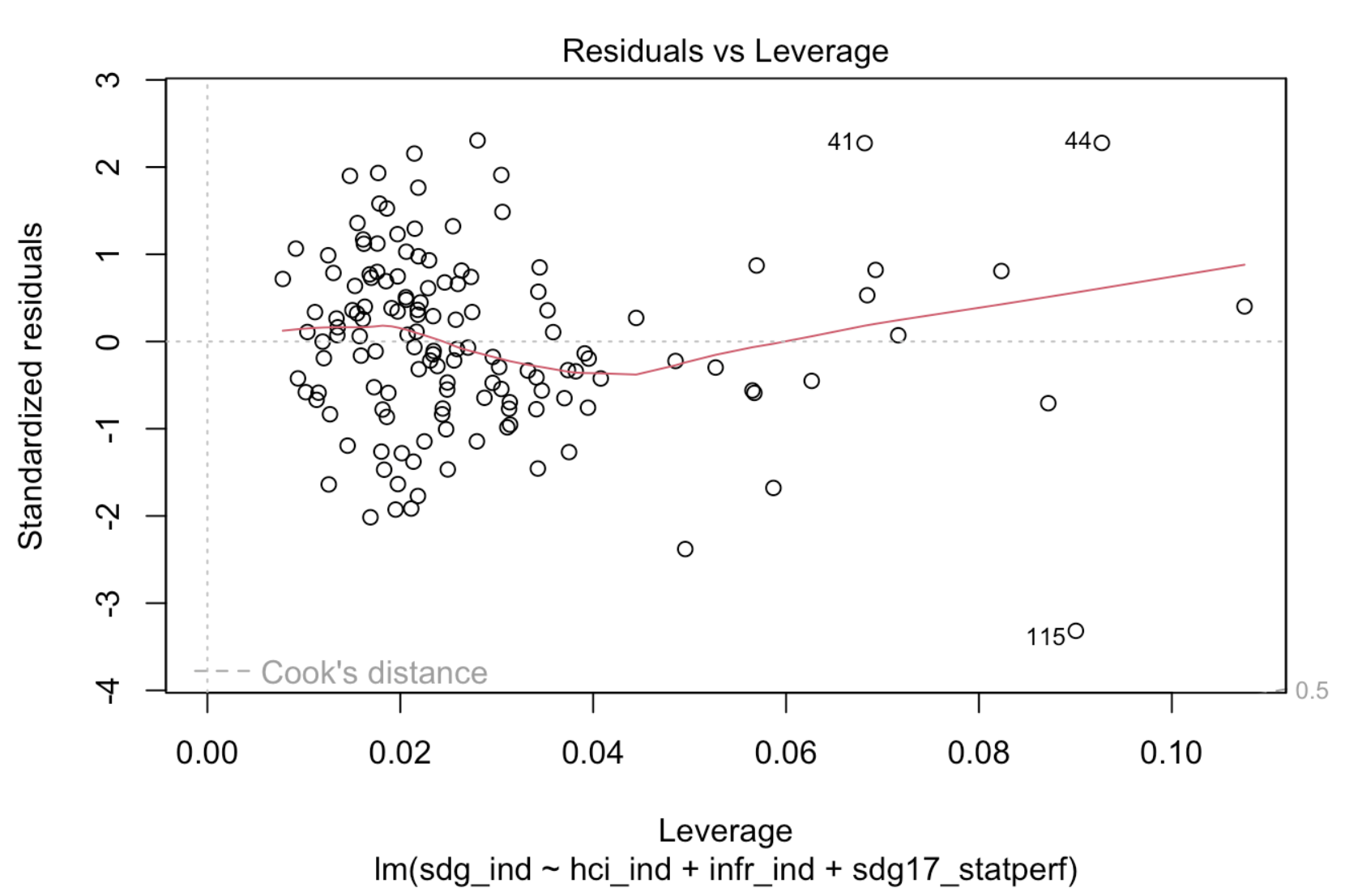
*2. Test for homoscedasticity*

In the below plot, we can also observe that the vertical spread of the residuals is equally distributed, which means the error term does not vary much as values of the outcome variable change. So, our model passes the test for homoscedasticity as well.



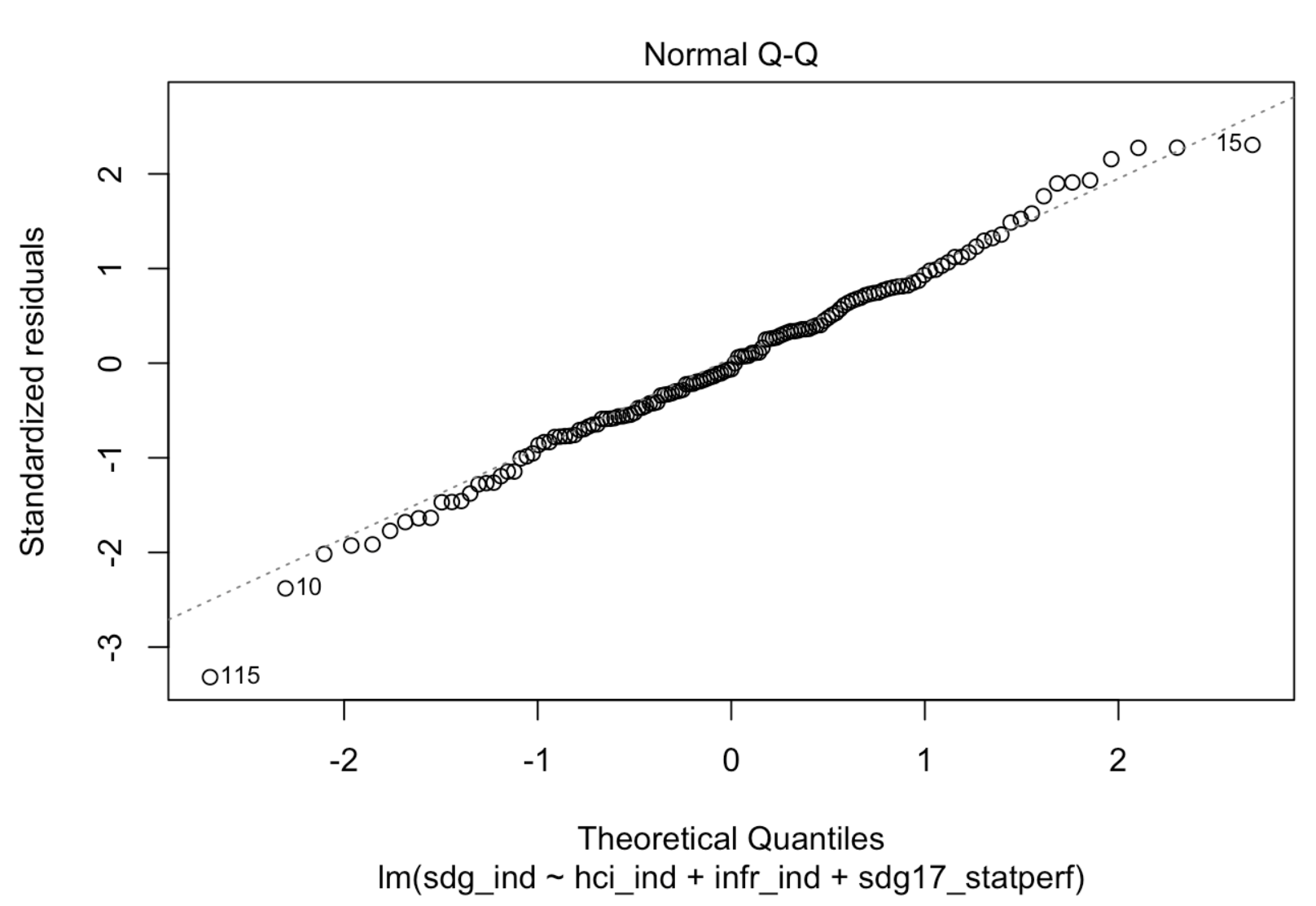
*3. Testing for Independence of residuals*

Based on the observations from the below “Residuals vs Leverage” plot, our model passes the test of independence of residuals as well. Large residual values on this plot would suggest that the model is not explaining some aspects of the data. Our model does not have any standardized residual values above 1. In R programming language, I double-checked and confirmed no observations with Cook’s distance value above 1.

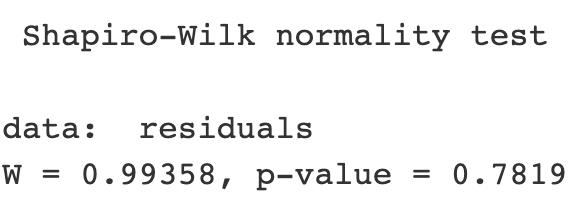


*4. Testing for Normality of the error distribution*

We can tell whether the error terms are normally distributed based on the observations from the below Q-Q plot. We want the residuals to be as close to the diagonal line as possible. However, generally, we rarely have real data where errors are perfectly normally distributed. So, some deviations are expected, and overall, it seems like our model passes the normality test. However, to double-check, I also apply the Shapiro-Wilk test.

****

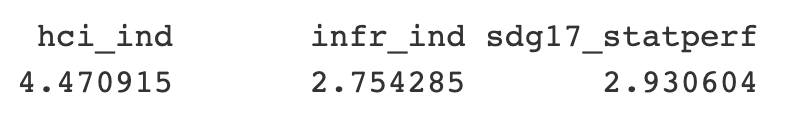
The null hypothesis for the Shapiro-Wilk test is that the data is normally distributed. In this case, the p-value for the Shapiro-Wilk test is way above the significance level, which means that we cannot reject the null hypothesis, and the data is normally distributed.



*5. VIF Score*

Last but not least, since we are dealing with a multiple linear regression model, we need to make sure there is no multicollinearity. So, we apply the VIF Score test. “A rough rule of thumb is that variance inflation factors [VIF] greater than 10 give some cause for concern.” (Vehklahti p.93) As we can see from the below table, the VIF scores for all three of our independent variables are below 5. These scores indicate some multicollinearity but are safely within an acceptable range.

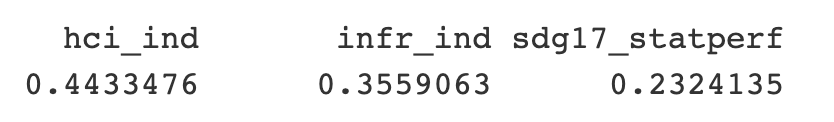
**VIF Scores:**



*Beta coefficient analysis*

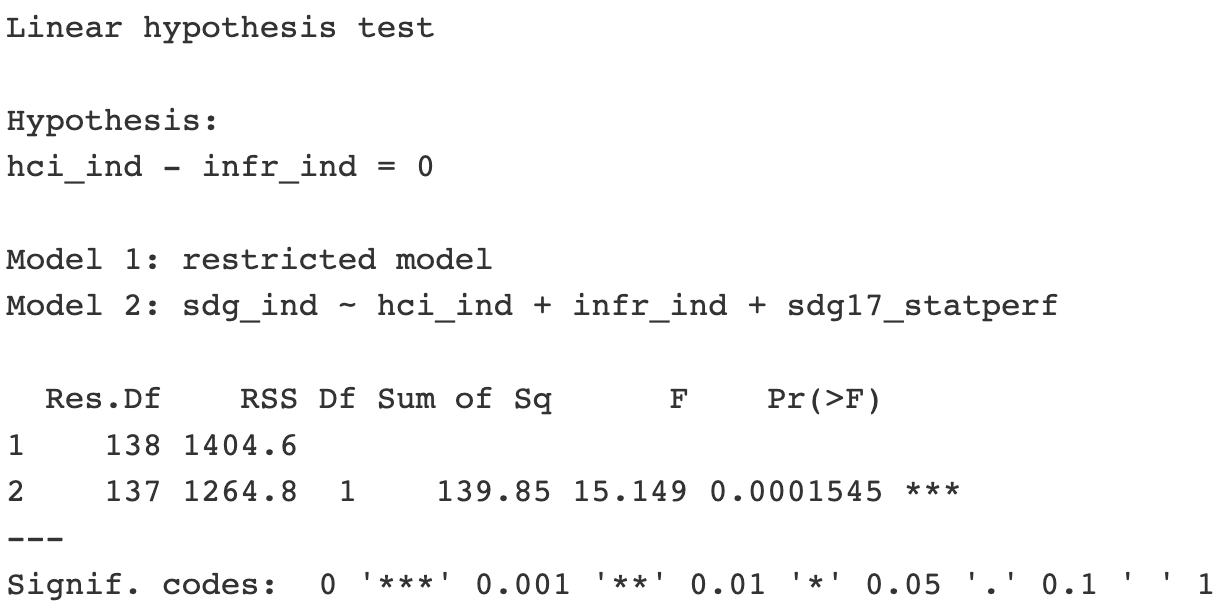
After we have confirmed that model meets all 5 assumptions of a multivariate regression model, we can proceed with the analysis of the model. In order to estimate the impact of each individual variable on the SDG Index, we can look at the beta coefficients. The standardized beta coefficients allow us to compare the effects of the variables on the same scale, regardless of the units of measurement. Below are the beta coefficients of our linear multivariate regression model. We notice that the beta coefficient for hci\_ind (Human Capital) is larger than the coefficient for infr\_ind (Infrastructure). This suggests that Human Capital has a stronger impact on the output variable, the SDG Index.

**Beta coefficients of the Multivariate Regression model**



However, we also need to make sure the difference between the two beta coefficients is statistically significant. I run the below linear hypothesis test, which is based on the null hypothesis that there is no difference between the effects of the two indices: hci\_ind and infr\_ind.

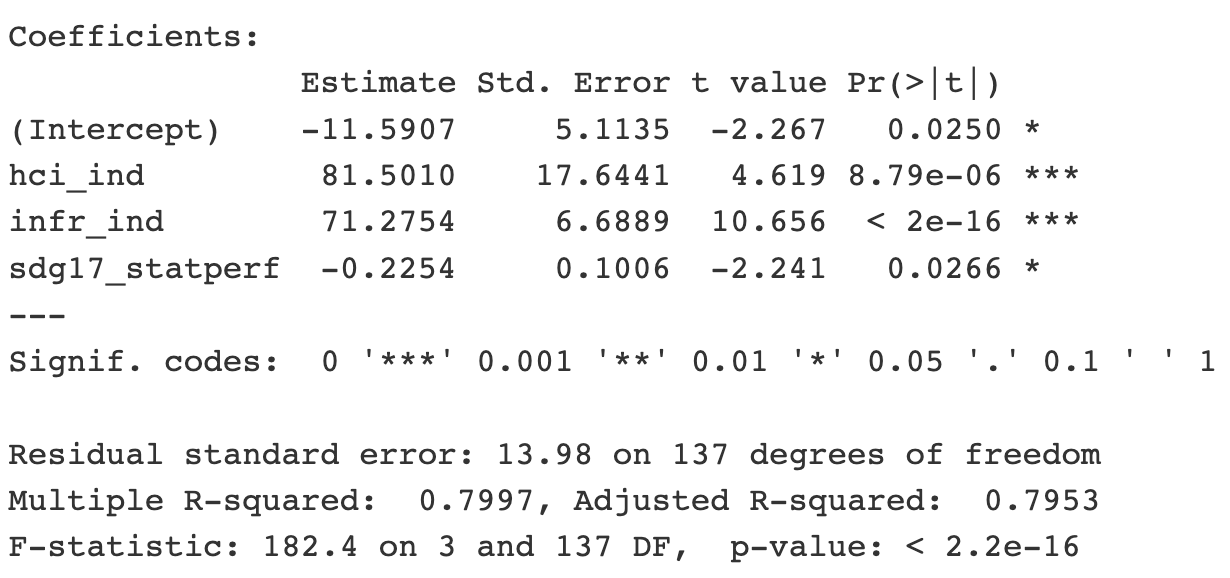
**Linear hypothesis test**



The associated p-value (Pr(>F) = 0.0001545) is far below 0.05, indicating strong evidence to reject the null hypothesis that the coefficients for hci\_ind (Human Capital Index) and infr\_ind (Infrastructure Index) are the same. So, the data provides strong evidence that the effect of Human Capital on the sdg\_ind (SDG Index) is different from the effect of Infrastructure (infr\_ind) on *sdg\_ind*.

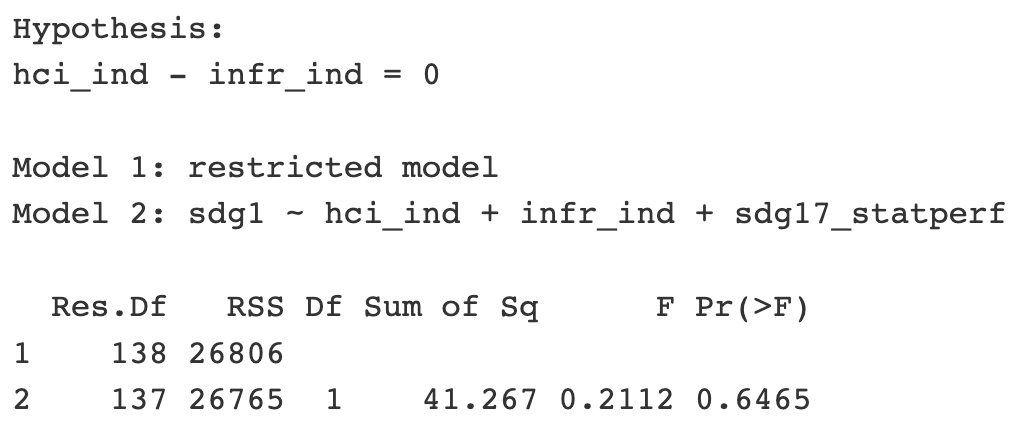
Next, I explore the relationship between Human Capital Index, Infrastructure Index and specific Sustainable Development Goals: SDG 1: No Poverty; SDG 5: Gender Equality; and SDG 13: Climate Action. I construct a multivariate multiple regression model with three left-hand variables, indicators for SDG 1, SDG 5, and SDG 13.

**Response SDG 1:**

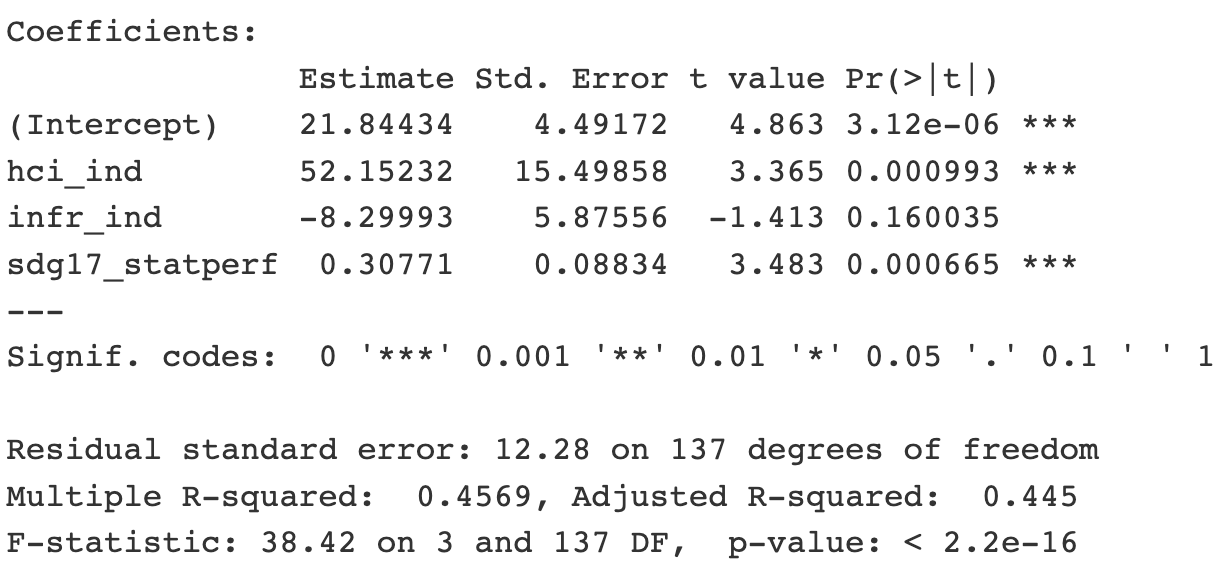


Based on the initial observation of the model summary, we can conclude that both human capital and infrastructure have a significant effect on poverty. However, we will need to explore further if there is a statistical difference between the effects of the two variables. Upon closer examination of the two beta coefficients, we find no statistically significant difference between the effects of the two explanatory variables.

**Linear hypothesis test**

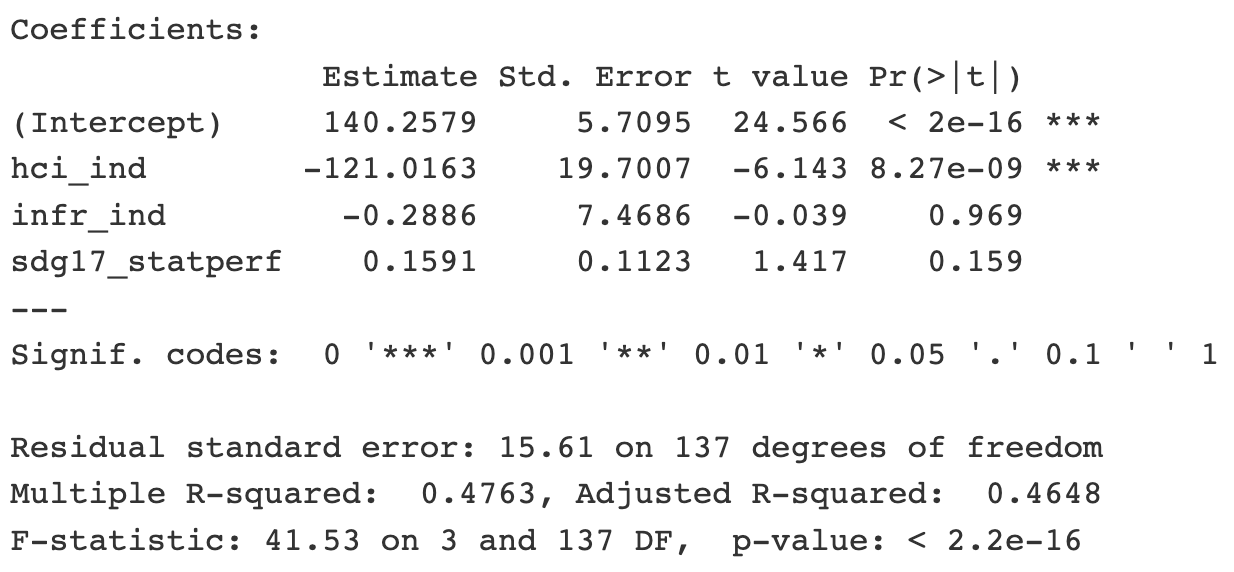


**Response SDG 5:**



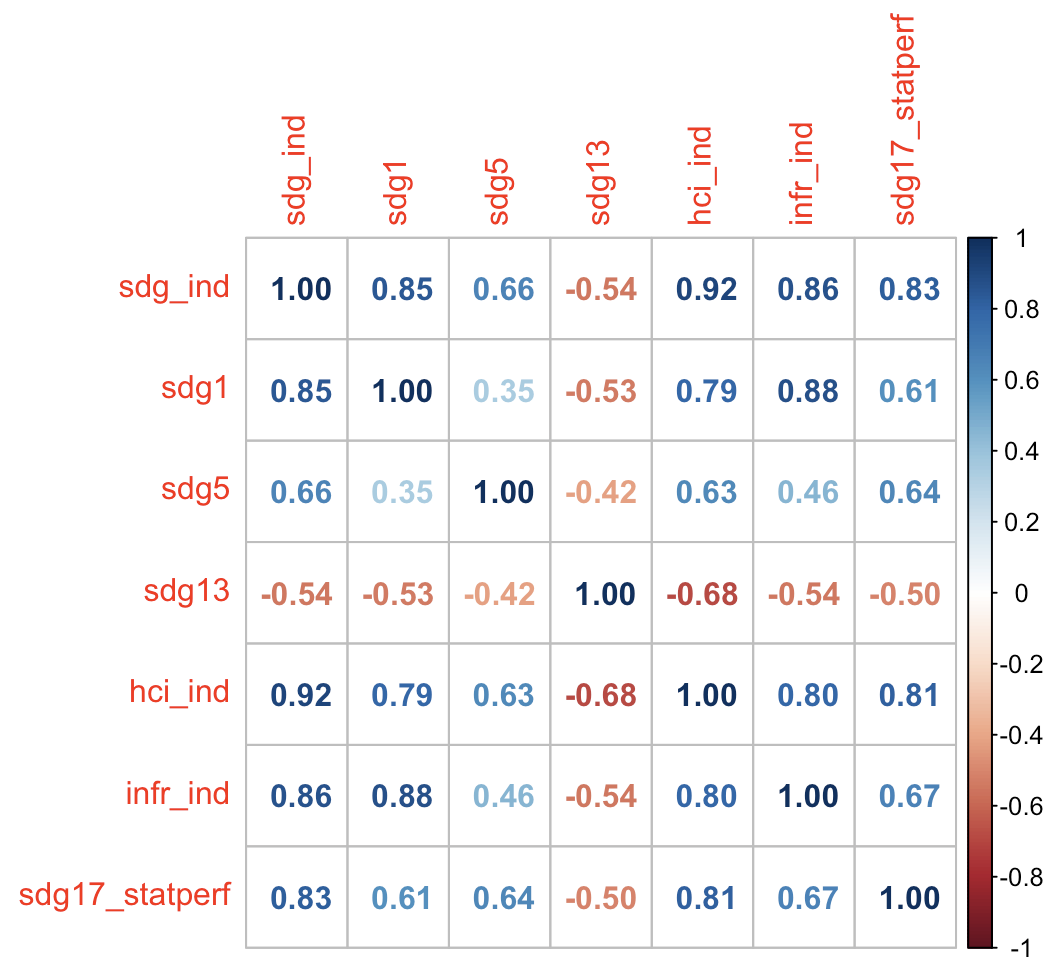
When we look at the response of SDG 5, we notice that Human Capital Index has a statistically significant impact on SDG 5, whereas Infrastructure Index does not. The value of the coefficient magnitude for hci\_ind (52.15) is also larger than the coefficient for infr\_ind (-8.30). Based on these observations, we can conclude that there is a statistical difference in the effects of human capital and infrastructure on SDG 5.

**Response SDG 13:**

****

The summary of the response to SDG 13 suggests that, once again, infrastructure does not have a statistically significant effect, but the impact of human capital is significant. So, we can claim that human capital has a statistically more significant effect on SDG 13 Climate action. However, we should also note that the coefficients are negative, which means there is a negative correlation between human capital and SDG 13. This is consistent with the basic correlations of the indicators in our dataset (please, see the Correlation Matrix table). I discuss these findings further in the conclusion.

**Correlation Matrix**



**IV. CONCLUSION**

Our findings confirm once again that both human capital and infrastructure are essential for the sustainable development of countries. They are both fundamentally important factors foretelling a country’s level of development. With that being said, based on our results, we can reject the Null Hypothesis that there is no statistical difference in the effect of human capital and infrastructure on a country’s SDG Index. The statistical analysis suggests a stronger inter-dependency between human capital and the SDG Index than with infrastructure. So, policy-makers facing the dilemma of choosing between investments in human capital and infrastructure should prioritize human capital if their goal is to advance the overall sustainable development agenda in the country.

However, we also found that Human Capital Index and the Infrastructure Index may have different levels of impact on specific objectives within the UN SDG framework. We discovered that human capital is a statistically significant indicator of a country’s performance on *SDG 5: Gender equality*, whereas infrastructure is not. We also established that while both indicators have a significant impact on a country’s performance on *SDG 1: No poverty*, there is no statistically significant difference between the effects of human capital and infrastructure on the poverty levels of a country. Last but not least, we figured that compared to infrastructure, there is a stronger inter-dependency between human capital and *SDG 13*: Climate action. However, there is a negative correlation between human capital and a country’s performance on climate indicators. This should not come as a surprise because developed countries with higher Human Capital Indexes produce far more carbon footprint than developing countries.[[11]](#footnote-11) It is another reminder that developed countries should transition to more sustainable solutions.

**V. WORKS CITED**

*Guterres urges countries to recommit to achieving SDGs by 2030 deadline*. (2023, April 25). UN News. https://news.un.org/en/story/2023/04/1136017

Johnson, J. B., & Joslyn R. A. (1991). Political Science Research Methods: Second Edition. Congressional Quarterly Inc.

Merriam-Webster. (n.d.). Infrastructure. In Merriam-Webster.com dictionary. https://www.merriam-webster.com/dictionary/infrastructure

Sachs, J.D., Lafortune, G., Kroll, C., Fuller, G., Woelm, F. (2022). From Crisis to Sustainable Development: the SDGs as Roadmap to 2030 and Beyond. Sustainable Development Report 2022. https://dashboards.sdgindex.org/downloads

The Investopedia Team. (2023, February 7). Infrastructure: Definition, Meaning, and Examples. Investopedia. <https://www.investopedia.com/terms/i/infrastructure.asp>

World Bank Group. (2023). The Human Capital Project: Frequently Asked Questions. In *World Bank*. https://www.worldbank.org/en/publication/human-capital/brief/the-human-capital-project-frequently-asked-questions

The World Bank Group. (2020, September 23). Data Catalog. Human Capital Index. https://datacatalog.worldbank.org/search/dataset/0038030

*The world’s top 1% of emitters produce over 1000 times more CO2 than the bottom 1% – Analysis - IEA*. (n.d.). International Energy Agency.

https://www.iea.org/commentaries/the-world-s-top-1-of-emitters-produce-over-1000-times-more-co2-than-the-bottom-1

United Nations. (n.d.). The 17 goals: Sustainable Development. United Nations. <https://sdgs.un.org/goals>

Vehkalahti, K., & Everitt, B. S. (2020). *Multivariate Analysis for the Behavioral Sciences: Second Edition.* CRC Press.

1. *Guterres urges countries to recommit to achieving SDGs by 2030 deadline*. (2023, April 25). UN News.  [↑](#footnote-ref-1)
2. Sachs, J.D., Lafortune, G., Kroll, C., Fuller, G., Woelm, F. (2022). From Crisis to Sustainable Development: the SDGs as Roadmap to 2030 and Beyond. Sustainable Development Report 2022. [↑](#footnote-ref-2)
3. United Nations. (n.d.). The 17 goals: Sustainable Development. United Nations. <https://sdgs.un.org/goals> [↑](#footnote-ref-3)
4. Ibid [↑](#footnote-ref-4)
5. The Investopedia Team. (2023, February 7). Infrastructure: Definition, Meaning, and Examples. Investopedia. [↑](#footnote-ref-5)
6. World Bank Group. (2023). The Human Capital Project: Frequently Asked Questions. [↑](#footnote-ref-6)
7. Merriam-Webster. (n.d.). Infrastructure. In Merriam-Webster.com dictionary. [↑](#footnote-ref-7)
8. United Nations. (n.d.). The 17 goals: Sustainable Development. United Nations. [↑](#footnote-ref-8)
9. Vehkalahti, K., & Everitt, B. S. (2020), p. 295 [↑](#footnote-ref-9)
10. Johnson, J. B., & Joslyn R. A. (1991), p. 319. [↑](#footnote-ref-10)
11. *The world’s top 1% of emitters produce over 1000 times more CO2 than the bottom 1% – Analysis - IEA*. (n.d.). International Energy Agency [↑](#footnote-ref-11)