Analysis of Correlations between Human Development Index and IQ scores

Olle Rehnfeldt, Bence Pal, Christoph Kurzbauer, Jakob Hutter

rehnfeldt_olle@student.ceu.edu, pal_bence@student.ceu.edu, kurzbauer_christoph@student.ceu.edu, hutter_jakob@student.ceu.edu

Central European University

7th of April 2024

1 INTRODUCTION

This analysis investigates the correlation between the Human Development Index (HDI), a measure of a country's achievements in the field of human development, and a country's average IQ score, a measure of an individual's analytic intelligence. The project examines in detail the variables interplay on a country and regional level utilizing univariate and bivariate techniques, with a focus on exploring the following hypotheses:

1. The average IQ and Human Development Index are positively correlated on a country level.
2. Less developed regions are more prone to outliers.
3. The IQ is biased towards western and more economically developed countries.

Analysis Design

The project aims to answer the hypotheses and test the statistical significance of findings, and is structured as follows:

1. Cleaning, preparation and merging
2. Univariate analysis for HDI and IQ data separately on country and regional level
3. Bivariate analysis for HDI and IQ

What the results can be used for

Firstly, the findings can be used to show that IQ is biased towards western and other economically developed countries, which could be used to readjust the measurement norm in a way to reduce the bias,

Secondly, legislators can use the result to evaluate their country's IQ performance relative to others with similar HDI scores or within their respective regions. They can further use these findings to readjust their strategies in terms of education and other human development indicators that could have an influence on IQ in the long-term.

Note on Causality

This project is an analysis of correlation, not causality. We believe that the causality is that some components of the HDI are having a causal mechanism toward countries' average IQ and are keen to highlight that we don't believe there is causality in the reversed direction.

2 DATA

The project combines two datasets, both datasets being from 2021. First, the average IQ score per country, the IQ score is a normed measure of the analytical intelligence of an individual. Second, the Human Development Index (HDI) per country, which is a measure of countries' achievements and standards in terms of health, education, and standard of living. The score ranges from one, being the highest achievable score, to zero being the lowest.

Variables

The two datasets where merged into one dataframe with the following columns (variables):

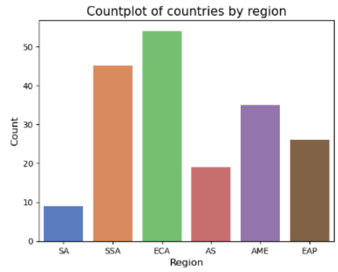
1. Country (Index)
2. Regions (referred to by abbreviations)

SA	South Asia
SSA	Sub Saharan Africa
ECA	Europe and Central Asia
AS	Rest of Asia
AME	Americas
EAP	East Asia and Pacific

3. HDI (Human Development Index 2021)

4. Average IQ (per country 2021)

Class Distribution



As this analysis aims to make conclusions on the regional level, we must examine the size of the different regions in terms of countries to know how insightful and representative these regions are and whether they have enough points to be statistically significant in later analyses. From the plot above one can observe that the most represented are the ‘Europe and Central Asia’, and ‘Sub-Saharan Africa’ regions. The ‘Americas’ region has 30 instances which is the rule of thumb for sufficient sample size while the other regions do not hit this mark.

Data Sources and Acquisition

The HDI dataset is sourced from the United Nations Development Program (UNDP), which is a UN development aid agency. The data gets collected and published yearly. A possible bias for this dataset could be towards countries which are able to collect more data easier, this could be due to financial advantages or large population size. Furthermore, the dataset does not consider data transparency, meaning some countries could omit or manipulate data willingly.

The IQ dataset is sourced from Kaggle and was published by Richard Lynn & Tatu Vanhanen, who publish yearly since 2002 IQ data as part of their book series “IQ and the wealth of Nations”. As mentioned in the introduction, we believe that a bias for the IQ towards Western and economically developed countries exists, as analytical thinking is deeply embedded in the educational culture, and the IQ is directed towards measuring exactly that.

For both IQ and HDI the available data for the year 2021 was taken for this project.

3 DATA CLEANING AND PREPARATION

We began by analyzing the Human Development Index (HDI) dataset, which comprises 206 rows and 1,008 columns, detailing the HDI of 194 countries from 1990 to 2021, including ranks, regions, and HDI categories (low, medium, high, very high). Additional details irrelevant to our study were omitted. We then curated a new data frame with the columns: 'country', 'hdicode', 'hdi_2021', and 'region', refining column names, verifying data types, and excluding countries with missing data for HDI. The original dataset’s regional HDI summaries were isolated and removed to focus on country-specific data, retaining countries lacking regional classification. After confirming that there are no duplicate entries and identifying columns with missing values, 43 countries without assigned regions—primarily from Europe, Asia, and North America—were allocated appropriate regions, including a unified region for the Americas, to avoid segregating the USA and Canada. Countries were then ranked by HDI and indexed by 'country', facilitating subsequent data merges.

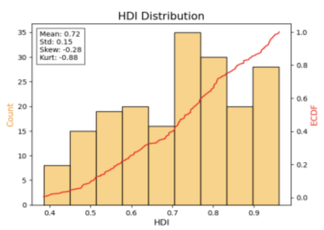
The IQ dataset, significantly tidier than the HDI data, underwent a similar process: selecting relevant columns, confirming the absence of duplicates, and organizing countries by IQ. Despite 41 countries having estimated IQs based on neighboring data, they were retained for analysis breadth. Like with the HDI data, the 'country' column was set as the index.

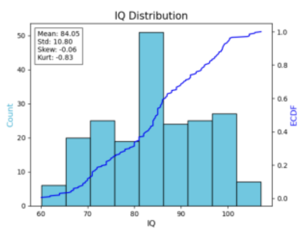
Both datasets were then merged via an outer join to compare country coverage, identifying and reconciling naming discrepancies. A final inner merge consolidated the harmonized datasets, ensuring only countries present in both were included in our analysis.

4 DESCRIPTIVE STATISTICS

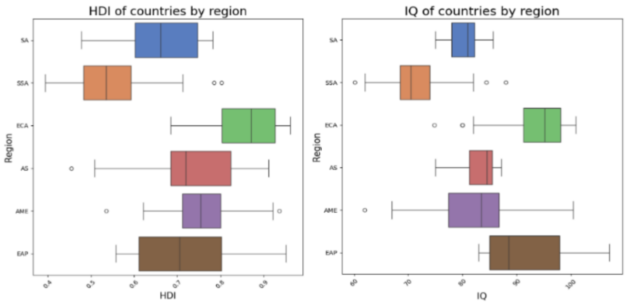
Univariate analysis

To initially understand the data distribution, histograms were generated, revealing the Human Development Index (HDI) is slightly left-skewed (skew statistic: -0.28), while the Intelligence Quotient (IQ) distribution is less skewed (skew statistic: -0.06), resembling a normal distribution more closely. Both HDI and IQ distributions display negative kurtosis (-0.83 and -0.91, respectively), indicating a higher concentration of values near the mean than expected in a normal distribution. Notably, although global IQ averages center around 100, our dataset shows a mean IQ of 84.05 per country, reflecting varying national populations.





Cumulative distribution analysis illustrates HDI values are evenly distributed across scores, whereas IQ values cluster more around the mean, evidenced by a steeper cumulative function slope.



Boxplot analysis identified outliers by comparing values outside 1.5 times the interquartile range, analyzed regionally to discern trends.

For HDI, five outliers were noted:

- In 'Sub-Saharan Africa', Mauritius and Seychelles exceeded others with HDIs of 0.802 and 0.782, respectively.
- Syria stood out in 'Rest of Asia' with a comparatively low HDI of 0.577.
- Haiti was a lower outlier in 'Americas' with an HDI of 0.535.
- The USA and Canada were notable outliers in the 'Americas' due to high HDIs of 0.921 and 0.936, respectively.

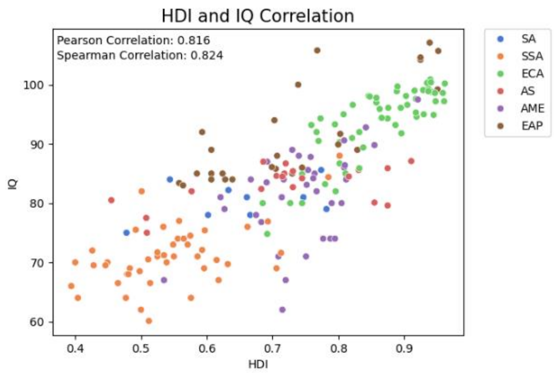
For IQ, six outliers were identified:

- Mauritius and Seychelles also had higher IQs of 88 and 84.4 in 'Sub-Saharan Africa', with Malawi being a lower outlier with an IQ of 60.1 in the region.
- Saint Lucia was a lower outlier in the 'Americas' with an IQ of 62.
- Kyrgyzstan was a lower outlier in 'Europe and Central Asia' with an IQ of 74.8, the countries of Turkmenistan, Tajikistan, Uzbekistan were also lower outliers, all with an average IQ of 80.

These findings, visualized in the histograms and boxplots, offer insights into regional trends and variations in HDI and IQ across countries. Notably, the 'Americas' region had upper outliers in the HDI region, but these were not repeated in the IQ region, whilst the Mauritius and the Seychelles were found outliers in both IQ and HDI statistics.

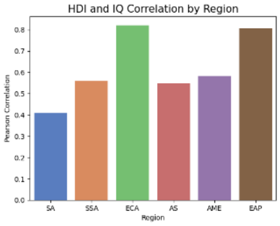
<u>Bivariate analysis</u>

In this analysis, we focused on the relationship between Intelligence Quotient and the Human Development Index, using a scatter plot to visualize their correlation, differentiated by region through color coding. The plot suggests a positive linear correlation between HDI and IQ, a finding supported by both Pearson and Spearman correlation coefficients. With a Pearson coefficient of 0.816, there's a strong linear relationship indicating that as HDI increases, so does IQ. The Spearman coefficient of 0.824 reinforces this, suggesting the relationship isn't significantly better described by a non-linear monotonic function, meaning that a linear model is an appropriate representation of the correlation between these variables.



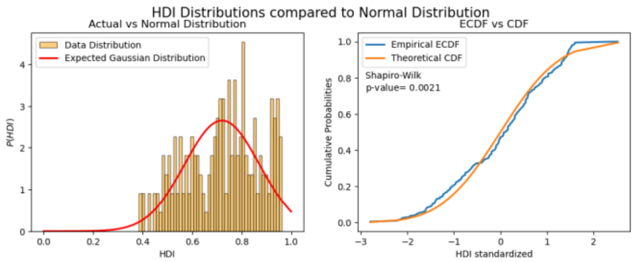
However, a closer examination reveals the relationship strength varies across the spectrum of HDI and IQ values; it's stronger where both are higher, while points near the origin are more dispersed. This discrepancy hints at regional variations in the correlation strength. Indeed, the scatter plots, segmented by region and color-coded, intuitively demonstrate differences in spread and correlation strength.

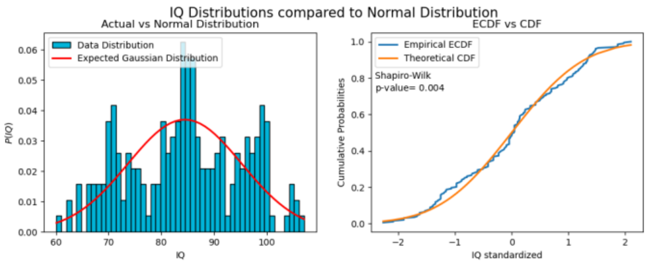
Further analysis involved creating individual scatter plots for each region, complemented by a bar chart comparing Pearson and Spearman correlations regionally. These results, detailed in another figure, which can be found in the notebook, highlight significant discrepancies in correlation across regions. Specifically, while 'East Asia and the Pacific' and 'Europe and Central Asia' show correlations similar to the overall dataset, South Asia, for instance, presents a markedly lower Pearson correlation of 0.410, pointing to substantial regional differences in how HDI and IQ interrelate.



5 FITTING DISTRIBUTIONS

The assessment of the HDI and ID distribution against a standard Gaussian distribution is portrayed in the accompanying graph, allowing for a detailed comparison between the empirical data and the theoretical model. Gaussian Distribution was chosen as the best expected fit after examining other potential distributions. The left histogram represents the actual distribution of HDI scores, with the overlaid red curve indicating the expected normal distribution. The discrepancy between the observed data and the Gaussian curve is quantified by the Shapiro-Wilk test, which tests the null hypothesis that the sample was drawn from a normal distribution, yielding a p-value of 0.0021, which suggests that we can reject the Null Hypothesis that the HDI distribution is drawn from a normal distribution on the 5% significance level. Complementarily, the graph on the left Cumulative Distribution graph illustrates the empirical cumulative distribution function (ECDF) against the theoretical cumulative distribution function (CDF) for HDI, where the proximity of the two curves up until the extremities of the distribution denotes a relative fit, albeit with some divergence, especially in the tails.

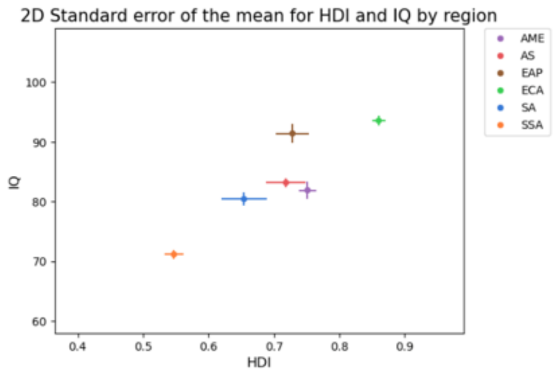




For the IQ we find similar results as the Shaprio-Wilk test returns a p-value of 0.004, so as before with HDI, we can reject the Null Hypothesis on the 5% level that the IQ distribution was drawn from a normal distribution.

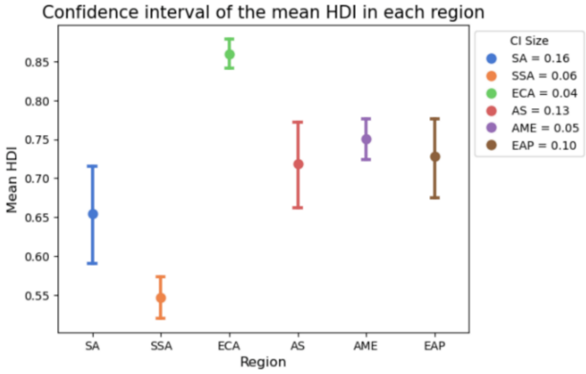
6 CONFIDENCE INTERVALLS AND ERROR BARS

In order to make a comparison on the regional level one has to define a point that represents the group. For similar clusters the two most common procedures are using the centroids, the artificial mean of the cluster, or the medoid, the data point closest to the center of the cluster. Here, the centroids of the different regions will be used and the standard error of the mean is plotted as error bars on top of the points. Also observed in the boxplots, one can see that the ‘Europe and Central Asia’ region is first in terms of both metrics, while the ‘East Asia and Pacific’ region would be second in terms of ranking as it has similar HDI levels but has significantly higher IQ than the ‘midfield’ consisting of ‘Rest of Asia’, ‘South Asia’, and the ‘Americas’ regions. While little significant difference can be observed for the ‘midfield’ the ‘Sub-Saharan Africa’ region is last by a huge margin in terms of both criteria.

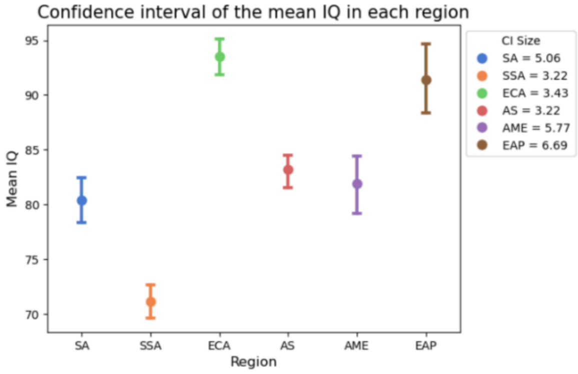


What could also be seen from the previous HDI vs IQ scatterplot is also showcased in this, namely that the ‘Europe and Central Asia’ and ‘Sub-Saharan Africa’ regions are more concentrated than the other regions and have smaller standard errors for the mean of the region. This and the differences in sample size, will lead to linear regression models fitting more to these regions and less to others when creating a model on the global scale.

Here, we visualize the 95% confidence interval of the mean HDI and IQ in each region, which helps us to understand the precision or uncertainty associated with our estimate of the mean HDI and IQ.



First, let's examine the mean HDI in each region. Four out of the six regions have their mean HDI between 0.65 and 0.75 and two are 'outliers’. These are the regions 'Europe and Central Asia' and 'Sub-Saharan Africa', the first one having a very high and the latter one having a very low mean HDI. These two, including the region 'Americas', have the smallest confidence intervals, which indicates that the estimate is more precise. The rest lies within the 0.65 and 0.75 mean HDI range, having



bigger confidence intervals, meaning that there is more uncertainty in the estimate. These regions are 'South Asia', 'Rest of Asia' and 'East Asia and Pacific'.

'We can see that, unlike the mean HDI, the mean IQ is more spread out, but the outliers stay the same. These are again 'Sub-Saharan Africa' and 'Europe and Central Asia', these are joined by 'East Asia and Pacific'. In the latter two regions, the mean IQ is significantly higher, and in 'Sub-Saharan Africa' it is significantly lower than in all the other regions. In the other three regions 'Asia', 'South Asia', and 'America' the mean IQ lies between 80 and 85. The latter two and 'East Asia and Pacific' have the biggest confidence intervals and thus the more uncertain estimate for mean IQ. The two outliers 'Europe and Central Asia', 'Asia', and 'Sub-Saharan Africa' have the smallest confidence intervals and more precise estimates for the mean IQ.

7 HYPOTHESIS TESTING

Getting quantifiable results on the similarity of means for the HDI and IQ scores of the regions can give us valuable insights into intra-regional similarities and differences. These can in turn deepen our understanding of trends that create biases in measurements and tell us which regions exhibit these biasing trends.

HDI

Initially we conduct an ANOVA test to investigate the null-hypothesis that all means are the same, meaning that if we can reject the null-hypothesis, we can conclude that there is at least one mean in the regions which is significantly different from the others. Performing the test on the HDI mean of every region yields a p-value of *1.08e-33*, which is sufficiently small to reject this null hypothesis at all reasonable significance levels.

Welsch t-test for HDI per region

Region Pair	p-value	size_pair	mean_pair	std_pair
AS vs EAP	0.8023	(19, 26)	(0.718, 0.728)	(0.131, 0.129)
AME vs EAP	0..425	(35, 26)	(0.751, 0.728)	(0.08, 0.129)
AS vs AME	0.3257	(19, 35)	(0.718, 0.751)	(0.131, 0.08)
SA vs AS	0.1764	(9, 19)	(0.654, 0.718)	(0.104, 0.131)
SA vs EAP	0.1013	(9, 26)	(0.654, 0.728)	(0.104, 0.129)
SA vs AME	0.0245	(9, 35)	(0.654, 0.751)	(0.104, 0.08)
SA vs SSA	0.0154	(9, 45)	(0.654, 0.547)	(0.104, 0.094)
SA vs ECA	0.0002	(9, 54)	(0.654, 0.86)	(0.104, 0.074)
ECA vs AS	0.0002	(54, 19)	(0.86, 0.718)	(0.074, 0.131)
SSA vs ECA	0.0	(45, 54)	(0.547, 0.86)	(0.094, 0.074)
SSA vs AS	0.0	(45, 19)	(0.547, 0.718)	(0.094, 0.131)
SSA vs AME	0.0	(45, 35)	(0.547, 0.751)	(0.094, 0.08)
SSA vs EAP	0.0	(45, 26)	(0.547, 0.728)	(0.094, 0.129)
ECA vs AME	0.0	(54, 35)	(0.86, 0.751)	(0.074, 0.08)
ECA vs EAP	0.0	(54, 26)	(0.86, 0.728)	(0.074, 0.129)

Furthermore, to get results on pair-wise similarities between the regions we utilize the t-test, testing the null hypothesis that two regional means are the same. To make conclusion about the validity of this test it is important to consider differences in variances between regions. When looking at the standard deviations, we see that they differ across the regions, by taking this into account, we find it most appropriate to use the Welch t-test, which assumes regions have different variances when comparing the means. Here we visualize the results in a table, including the sample sizes, standard deviations and means of the regions, each p-value has been rounded with four significant digits.

The results found is that there seems to be three regions which the test indicates are all similar, these are 'Rest of Asia', *'*Americas*'*, and 'East Asia Pacific'. With their high p-values we cannot reject the null hypothesis that they share mean HDI. Although there are other pairs which also have p-values which, depending on the significance level chosen, could be large enough to reject the null hypothesis, these all contain the 'South Asia' region which has a very small sample size, hence these results should be taken with a great grain of skepsis to their actual validity.

IQ

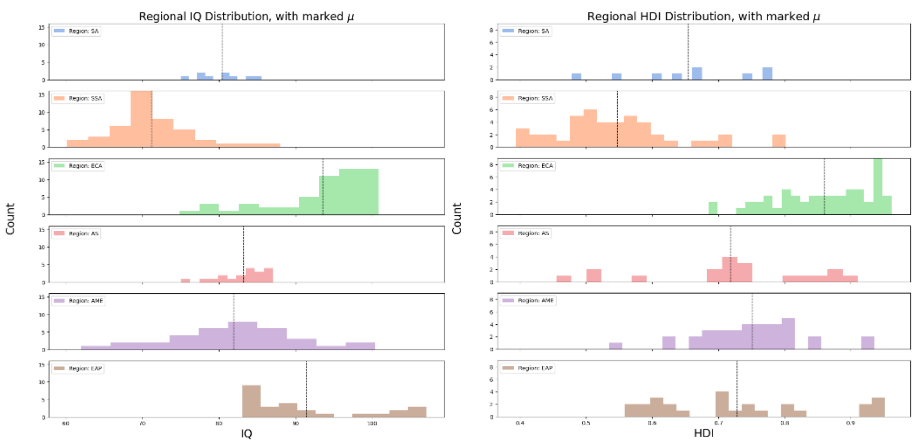
Continuing with the same procedure as for HDI, we perform the ANOVA-test on the regions to investigate mean IQ similarity, yielding a p-value of *4.81e-39*, meaning we can again reject the null hypothesis that the mean IQ is the same across all regions.

Welsch t-test for IQ per region

Region Pair	p-value	size_pair	mean_pair	std_pair
AS vs AME	0.4337	(19, 35)	(83.174, 81.9)	(3.344, 8.397)
SA vs AME	0.4157	(9, 35)	(80.422, 81.9)	(3.29, 8.397)
ECA vs EAP	0.2481	(54, 26)	(93.556, 91.404)	(6.279, 8.283)
SA vs AS	0.0565	(9, 19)	(80.422, 83.174)	(3.29, 3.344)
AS vs EAP	0.0001	(19, 26)	(83.174, 91.404)	(3.344, 8.283)
SA vs SSA	0.0	(9, 45)	(80.422, 71.167)	(3.29, 5.361)
SA vs ECA	0.0	(9, 54)	(80.422, 93.556)	(3.29, 6.279)
SA vs EAP	0.0	(9, 26)	(80.422, 91.404)	(3.29, 8.283)
SSA vs ECA	0.0	(45, 54)	(71.167, 93.556)	(5.361, 6.279)
SSA vs AS	0.0	(45, 19)	(71.167, 83.174)	(5.361, 3.344)
SSA vs AME	0.0	(45, 35)	(71.167, 81.9)	(5.361, 8.397)
SSA vs EAP	0.0	(45, 26)	(71.167, 91.404)	(5.361, 8.283)
ECA vs AS	0.0	(54, 19)	(93.556, 83.174)	(6.279, 3.344)
ECA vs AME	0.0	(54, 35)	(93.556, 81.9)	(6.279, 8.397)
AME vs EAP	0.0	(35, 26)	(81.9, 91.404)	(8.397, 8.283)

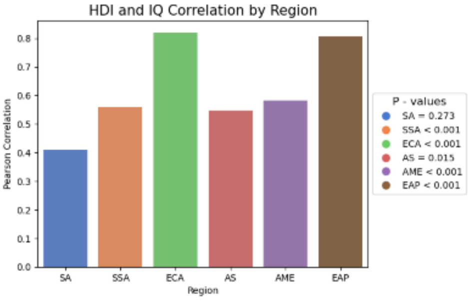
To see the pair-wise similarities we again use the Welch t-test after observing varying sizes of standard deviations, the results are illustrated with the same table structure as above. We can conclude that the results indicate a high similarity between the IQ for region pairs: *(*'Rest of Asia', *'*Americas*')* and ('Europe Central Asia', 'East Asia Pacific'*)*. ('South of Asia', *'*Americas*')* also exhibit similarities but due to the low sample size these should again be interpreted very conservatively.

It is important to further emphasize the risk of running into type 3 errors – rejecting the null hypothesis for the wrong reason – when using the small sample sizes, we do here. Moreover, one should also consider that the ANOVA-test and Welch t-test require the data it compares to be normally distributed. When plotting the distributions (figure below), the data does look somewhat normally distributed for the different regions, but to varying extents. Taking this into consideration, the very low p-values of the ANOVA-test for HDI and IQ also indicate that the means seem to vary to some extent from region to region. As has been demonstrated, the regions 'Rest of Asia' , 'Americas', and 'East Asia Pacific' for HDI and the region pairs of ('Rest of Asia', 'Americas') and, ('Europe Central Asia', 'East Asia Pacific') for IQ, while accounting for limitations to seem to be similar.

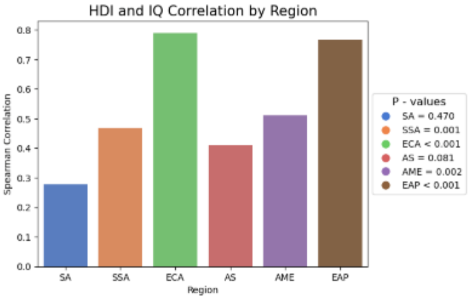


8 STATISTICAL SIGNIFICANCE TEST

Extending our initial bivariate analysis, we test the statistical significance of the Pearson and Spearman correlation coefficient in each region and are calculating the respective p-value. The correlation was calculated between the HDI and IQ in each region. The focus will lie on the p-value which indicates the probability of observing a correlation coefficient as extreme as, or more extreme than, the one calculated from the sample data, assuming that the correlation is zero. Note that the p-values displayed on the plots are rounded to the 3rd digit after the comma.



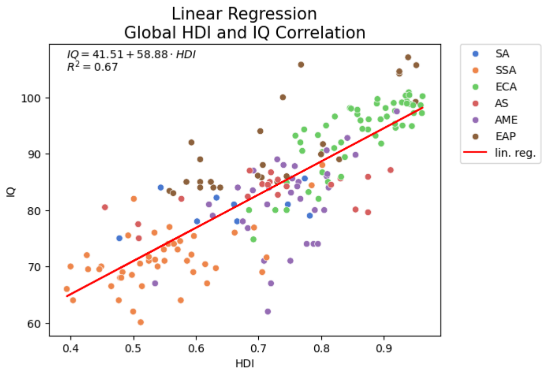
The Pearson correlation coefficient measures the linear relationship between the HDI and IQ. The regions 'Sub-Saharan Africa', 'Europe and Central Asia', 'Americas', and 'East Asia and Pacific' have very low p-values and are well below the 1% significance level. This indicates that the observed linear correlation between HDI and IQ in these regions is unlikely to have occurred by random chance if the null hypothesis, that there is no correlation, was true. This is not the case for the 'Rest of Asia' region for the 1% level but not for the 5% significance level. For the region 'South Asia' we obtain a relatively high p-value at approximately 0.273, which is well above the 5% and even 10% significance levels. Thus, we do not have enough statistical evidence to reject the null hypothesis, that there is no correlation between HDI and IQ. So, there is not enough evidence to support a significant linear correlation between the variables in the 'South Asia' region.



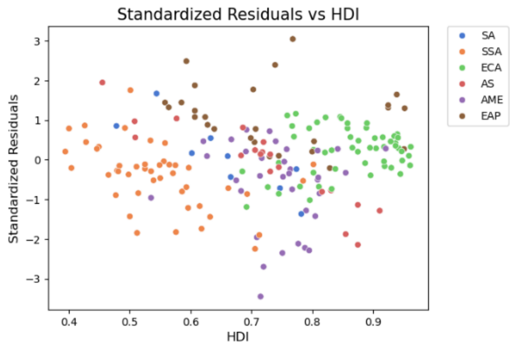
The Spearman correlation coefficient measures the non-linear monotonic relationship between the HDI and IQ. Here, two regions have a p-value significantly lower than 0.01, so there is a significant non-linear correlation between the two variables in these two regions on the 1% level. These regions are 'Europe and Central Asia' and 'East Asia and Pacific'. The regions 'Sub-Saharan Africa' and 'America' have p-values of 0.001 and 0.002 respectively, which are also lower than 0.01, so we can conclude that in these regions there is also a significant non-linear correlation between HDI and IQ. Lastly, we have the region 'Rest of Asia' with a p-value of 0.081 and 'South Asia' with a p-value of 0.470. Just like for the Pearson correlation, we do not have enough statistical evidence to conclude that there is a statistically significant correlation between the variables in 'South Asia'. For the region of 'Rest of Asia', we could only reject the null hypothesis, that there is no correlation between HDI and IQ, at the 10% significance level.

9 LINEAR REGRESSION

Our study utilized linear regression to analyze the correlation between the HDI and IQ, as depicted in the beneath figure with a regression equation of IQ = 41.51 + 58.88 × HDI. The intercept indicates a theoretical baseline IQ of 41.51 in the absence of HDI, while the slope suggests that each one-point increase in HDI corresponds to a 58.88 increase in IQ. The R^2 value of 0.67 indicates that 67% of the variation in IQ is explainable by HDI across the countries analyzed.



The regression analysis further reveals regional differences; data points are densely clustered at higher HDI and IQ levels but more scattered at lower levels. This pattern suggests varying strengths of the HDI-IQ relationship across development spectrums, which warrants a deeper examination on a regional basis to uncover the complexities within the global context of HDI and IQ interplay.When we look at the residuals on the global scale the plot looks fine and does not imply any possible biases. To give a more precise metric, about 94.15% of the standardized residuals are within 2 standard deviations to the mean of 0.



If we look at the same plot with the different regions highlighted, we can see that the domination of the 'ECA' and 'SSA' region in terms of data instances and spread lead to serious biases when trying to predict the values of some regions using the previous linear regression model. The regression fitted well to the 'ECA' and 'SSA' regions but there is a systematic underestimation of the 'EAP' region and a systematic underestimation of the 'AME' region. This means that our model would not generalize well for predicting the metrics in these regions. This is an important factor to consider if one tries to use the model on regional level.

10 DISCUSSION AND PRELIMINARY CONCLUSIONS

The findings from our analysis provide evidence in support our hypotheses concerning HDI and IQ:

1. The average IQ and Human Development Index are indeed positively correlated on a country level.
2. IQ and HDI Distribtution are not drawn from an expected normal distribution.
3. Less developed regions are indeed more prone to outliers.
4. The higher correlations between HDI and IQ observed in western regions suggest that the IQ measure of intelligence may indeed be more reflective of the educational and cognitive styles predominant in Western societies, and is therefore and indicator for bias.
5. IQ and HDI are significantly different for the regions by the ANOVA. 'Rest of Asia', 'Americas', and 'East Asia Pacific' as well as the region pairs ('Rest of Asia', 'Americas') and, ('Europe Central Asia', 'East Asia Pacific') have similar means.
6. Linear Regression is a good predictor in bivariate analysis for ECA, SSA and SA regions.

The implications of these findings are manifold. For policymakers and educators, the correlation between HDI and IQ underscores the importance development strategies that encompass not only economic growth but also improvements in health, education, and living standards. For researchers, the observed biases and outliers invite further investigation of IQ biases and the development of more culturally sensitive measures of cognitive ability.

In conclusion, this analysis not only supports intial hypothesized findings between Human Development and Analytic Inteligence but also prompts critical reflections on the measures we use to assess these constructs.