Comparing World Quality of Life Measures: Parametric vs. Nonparametric Approaches

Katherine M. Prioli

December 20, 2018

Abstract

BACKGROUND Increasing globalization has generated interest in comparing countries by key quality of life (QoL) measures such as life expectancy, gender equality, and happiness, among others. When considering how countries compare by QoL, it is vital to understand how the measures are related, if at all, and in what ways the statistical tests chosen affect the results. METHODS A dataset containing country names and selected QoL measures for 2016 was explored through descriptive statistics and sensible univariate and bivariate visualizations. Five sets of hypotheses pertaining to relationships within the data were generated and tested with both nonparametric methods and their parametric equivalents, and results of these parallel analyses were compared. RESULTS Nonparametric methods dominated, performing acceptably for all five cases, whereas parametric methods were acceptable in only two of the five cases. CONCLUSION Nonparametric methods represent a powerful, important toolset for use in analyzing data that fails the assumptions needed for traditional parametric methods.

Background

With increasing economic globalization, a natural topic of interest is how the world's nations compare with respect to quality of life (QoL). Several organizations monitor global QoL indicators and report single-dimension or aggregate values for indicators of interest. For example, the World Bank reports Gross Domestic Product (GDP), which is a single-dimension indicator often strongly predictive of QoL in a given country (The World Bank (2018)). Additionally, the World Health Organization reports infant mortality rate, life expectancy at birth, and life expectancy at 60 years of age (World Health Organization (2018b), World Health Organization (2018a)).

Other quality of life measures represent compound scores or indices based on several inputs. For example, the United Nations calculates an annual Human Development Index (HDI), representing the developmental level of each country on a scale of zero to one based on several factors, including life expectancy at birth, years of schooling, and per-capita income (The United Nations Development Programme (2018b)). The HDI also categorizes countries into four levels of development (low, medium, high, and very high). Similarly, the Social Progress Imperative publishes the Social Progress Index (SPI), ranging from 0 to 100, and comprising over 50 dimensions in three broad categories: basic human needs (e.g., nutrition, safety), foundations of wellbeing (e.g., basic knowledge, environmental quality), and opportunity (e.g., personal rights, freedoms) (Social Progress Imperative (2018b)). The World Economic Forum's Global Gender Gap Index reports a gender equality index, scaled from 0-1, based on measurements of gender-related gaps in such dimensions as economic participation, level of education, health and survival, and political offices held (World Economic Forum (2016b)). Finally, the World Happiness Report calculates a score from 0-10 by considering per-capita GDP, healthy life expectancy, social support, freedoms, and perception of corruption, among others (Helliwell, Layard, and Sachs (2018)).

The objective of this analysis is to explore the distributions of and relationships between key QoL indicators using both nonparametric and parametric methods, and to assess the appropriateness of each method used.

Methods

The dataset used in this analysis, titled `alldata`, was generated for the MAT 8790 course (Prioli (2018b)). It consists of country-level variables for calendar year 2016 as described in Table 1.

Table 1. `alldata` dataframe contents.

Source	Variable Name	Description
`countrycode` package	`country`	Country names
Social Progress Imperative (2018a)	`SPI`	Social Progress Index value (scale of 0:100)
The World Bank (2018)	`logGDP`	Gross Domestic Product, log transform (valued in $US 2018)
The United Nations Development Programme (2018a)	`HDIrank`	Human Development Index ranking

Source	Variable Name	Description
The United Nations Development Programme (2018a)	`HDIindex`	HDI index value (scale of 0:1)
The United Nations Development Programme (2018a)	`HDI_cat`	HDI index category (5 levels)
Helliwell, Layard, and Sachs (2018)	`happiness`	World Happiness Score (scale of 0:10)
World Economic Forum (2016a)	`gendereq`	Gender Equality Index (scale of 0:1)
World Health Organization (2018b)	`infantmort`	Infant mortality rate
World Health Organization (2018a)	`birth_MF`	Life expectancy at birth, males & females
World Health Organization (2018a)	`sixty_MF`	Life expectancy at 60 years, males & females

All variables pertain to the calendar year 2016. Missing values were omitted from the dataset to ensure that the tests of interest could be performed.

For each variable except `country`, descriptive statistics were run and a sensible visualization was generated, following which a correlation matrix was produced to examine pairwise relationships between continuous variables.

Nonparametric and Parametric Analyses

Based on the data exploration results, five sets of formal hypotheses were generated about the data (Table 2), and sensible nonparametric tests and their parametric equivalents were chosen to test these hypotheses.

Table 2. Analyses Performed

Analysis	Variable(s)	Null Hypothesis	Alternative Hypothesis	Nonparametric Test	Parametric Test
1	HDIindex	The sample median is equal to its mean	The sample median differs from its mean	One-Sample Sign Test	One-Sample t-Test
2	HDIindex, SPI	Human development and social progress are not associated	Human development and social progress are correlated	Kendall's Tau	Pearson's Correlation Test
3	logGDP, infantmort	There is no relationship between log(GDP) and infant mortality	There is a relationship between log(GDP) and infant mortality	Hoeffding's Test	Pearson's Correlation Test
4	happiness	Happiness is normally distributed	Happiness is not normally distributed	One-Sample Kolmogorov-Smirnov Test	Shapiro-Wilk Test
5	HDI_cat, infantmort	Infant mortality rate is the same across levels of human development	Infant mortality rate differs by level of human development	Permutation F-Test	ANOVA

A two-sided one-sample test was chosen for Analysis #1 because the `HDIindex` distribution is very non-normal, yet its median and mean appear quite similar and the standard deviation is small. The one-sample sign test assumes that the sample is random with independent draws, and the data are continuous. Its parametric equivalent, the one-sample t-test, shares these assumptions and also requires normality of the sampling distribution. The formal hypotheses tested for the sign test were $H_0 : m =$ `mean(alldata$HDIindex)` versus $H_A : m \neq$ `mean(alldata$HDIindex)`, whereas the one-sample t-test assessed the complementary hypotheses $H_0 : \mu =$ `median(alldata$HDIindex)` versus $H_A : \mu \neq$ `median(alldata$HDIindex)`. The sign test was carried out using `BSDA::SIGN.test()` and the one-sample t-test via `t.test()`.

Analysis #2 was motivated by the correlation matrix analysis, which indicates a linear relationship is likely between `SPI` and `HDIindx`. Kendall's Tau was chosen as the nonparametric test, with the usual Pearson's test as the parametric alternative. Both tests assume continuous data and are equipped to detect linear dependence, and employed two-sided alternatives. Additionally, because simulation is the preferred method when ties are present for Kendall's Tau, the variables were first assessed for ties before selecting a method for Kendall's Tau. The competing hypotheses for this analysis were $H_0 : \tau = 0$ vs. $H_A : \tau \neq 0$ (for Kendall's Tau) and $H_0 : \rho = 0$ vs. $H_A : \rho \neq 0$ (for Pearson's Test). Both tests were carried out via `cor.test()` using `method = "kendall"` or `method = "pearson"` as appropriate.

Correlation tests were chosen for Analysis #3 to determine whether increasing GDP is correlated with decreasing infant mortality. Hoeffding's Test was selected because it is sensitive to any departure from independence. Because the relationship appears to be roughly nonlinear, Pearson's test was chosen as the parametric comparator. Both tests employed two-sided alternatives and assume continuous data, and Pearson's test additionally assumes a linear relationship. For Hoeffding's test, the formal hypotheses tested were $H_0 : \Delta = 0$ vs. $H_A : \Delta \neq 0$ where

$$\Delta = \int_{X \in x} \int_{Y \in y} [F_{XY}(x, y) - F_X(x)F_Y(y)]^2 dF_{XY}(x, y).$$

For Pearson's test, the competing hypotheses were $H_0 : \rho = 0$ versus $H_A : \rho \neq 0$. These tests were carried out using `testforDEP::testforDEP()` with the options `test = "HOEFFD"` and `test = "PEARSON"` respectively.

Analysis #4 was chosen because `happiness` appears to have a symmetric, possibly normal distribution on univariate analysis. Both the one-sample Kolmogorov-Smirnov test and the Shapiro-Wilk test assume a continuous sample distribution. The formal hypotheses for this analysis were H_0 : `happiness` is normally distributed vs. for H_A : `happiness` is not normally distributed. Kolmogorov-Smirnov testing was performed via `ks.test()` using the arguments `"pnorm", mean(alldata$happiness), sd(alldata$happiness))` to specify the continuous comparator normal distribution having mean and standard deviation as found in `happiness`. Shapiro-Wilk testing was carried out using `shapiro.test()`.

To test for a difference in infant mortality rate by human development category, Analysis #5 employed the permutation F-test and ANOVA as the nonparametric and parametric tests respectively. Both tests assume samples are independent, and ANOVA additionally assumes a normal distribution with contstant variance. The competing hypotheses were H_0 : the cumulative distribution functions (CDFs) for infant mortality were equal across all four human development levels vs. H_A : at least one of these CDFs differs. The permutation F-test was performed via `jmuOutlier::perm.f.test()` with 1,000 simulations because the sample size was prohibitively large for calculating an exact p-value. The `aov()` function was used for ANOVA testing.

Results of these five paired analyses were compared in the context of the data and assumptions needed. All tests were performed at level $\alpha = 0.05$ in R. The dataset, full code, and this report are available in an online repository (Prioli (2018a)).

Results

Descriptive Statistics and Visualizations

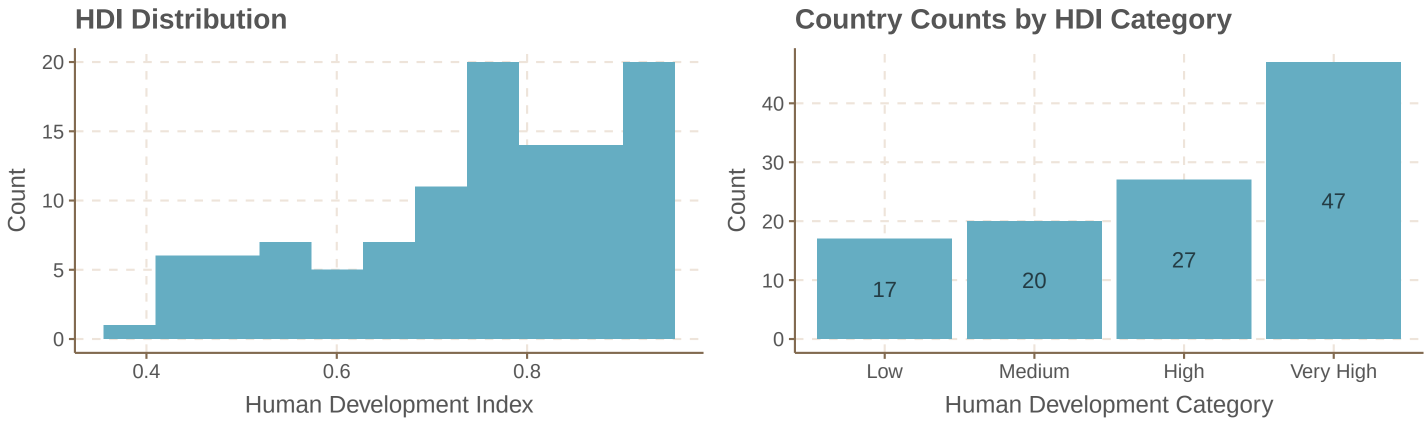
The original dataset represented 205 countries, and the analytic dataset contained 111 countries after omitting those with missing values. Of the 94 countries omitted, 26 were highly developed by HDI category, 20 were in the low developmental category, and 18 had unknown HDI category (Table 3).

Table 3. Omitted Countries by HDI Category

HDI Category	Frequency
Low	20
Medium	18
High	26
Very High	12
NA	18

Descriptive statistics and univariate visualizations are presented in Figs. 1-8. Figure 1 depicts the distribution and descriptive statistics of the `HDIindex` variable, along with counts by country for the `HDI_cat` variable. `HDIindex` has a very nonnormal appearance, yet has similar mean and median, and a small standard deviation as compared to the mean. The highest category of human development is the most represented in the data.

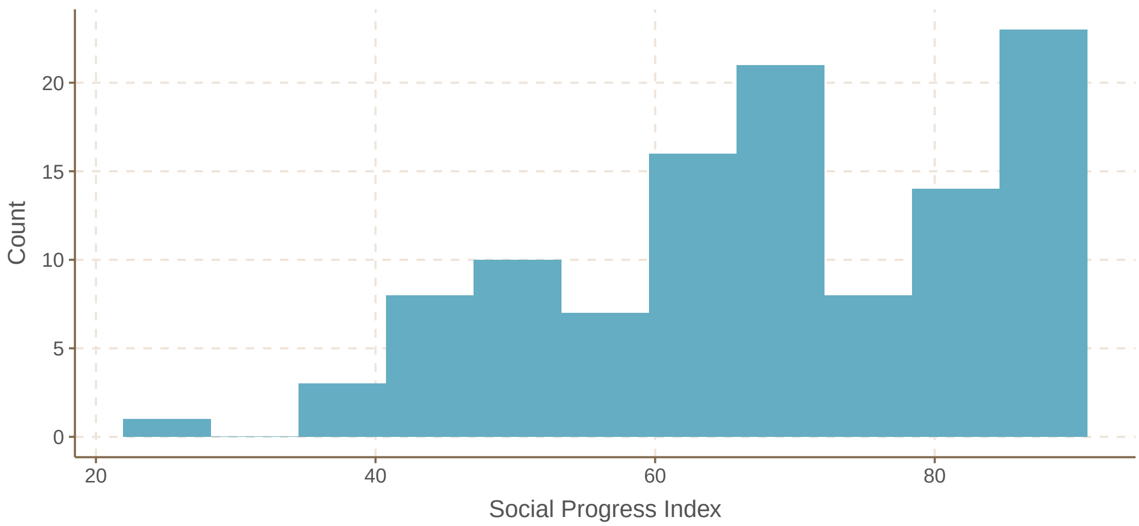
Figure 1. Human Development Index



Min	Q1	Median	Mean	Q3	Max	SD
0.405	0.653	0.772	0.745	0.867	0.951	0.15

`SPI`, described in Fig. 2, appears to have a multimodal distribution, with mean and median quite similar in value.

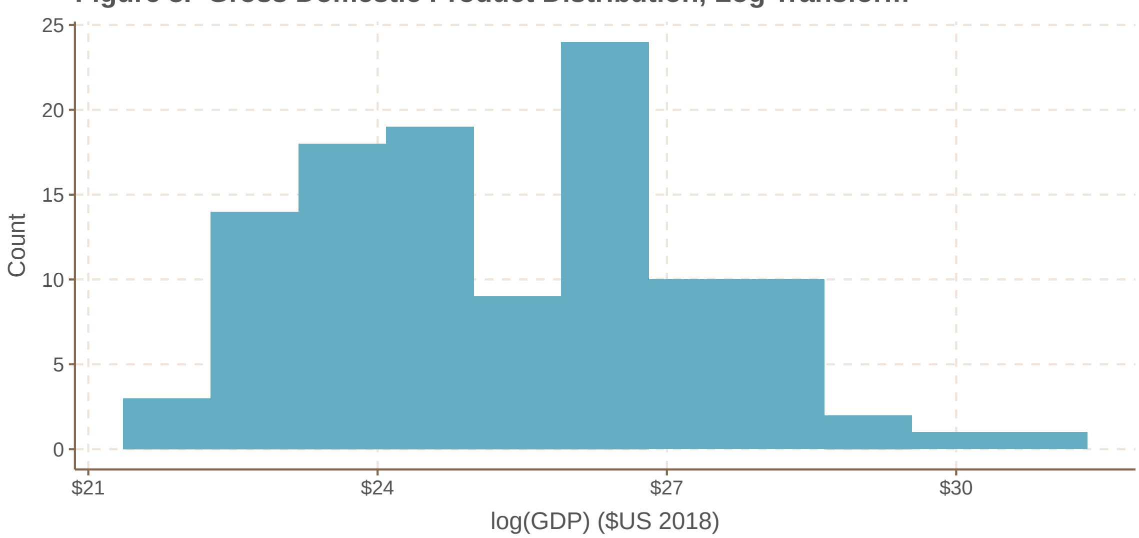
Figure 2. Social Progress Index Distribution



Min	26.92
Q1	58.42
Median	68.94
Mean	68.624
Q3	82.37
Max	89.62
SD	15.381

The distribution of `logGDP` (Fig. 3) appears possibly bimodal with mean and median reasonably close in value and standard deviation smaller than both.

Figure 3. Gross Domestic Product Distribution, Log Transform



Min	21.466
Q1	23.73
Median	25.014
Mean	25.323
Q3	26.666
Max	30.555
SD	1.942

Per Fig. 4, the World Happiness Score appears roughly normally distributed.

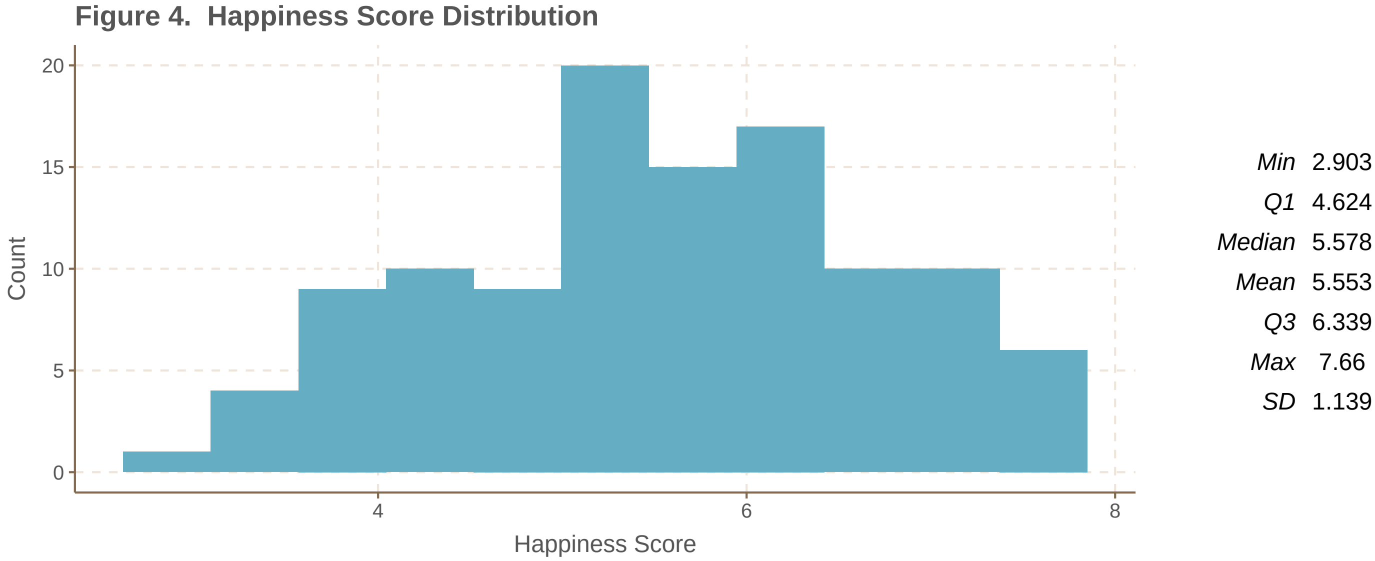
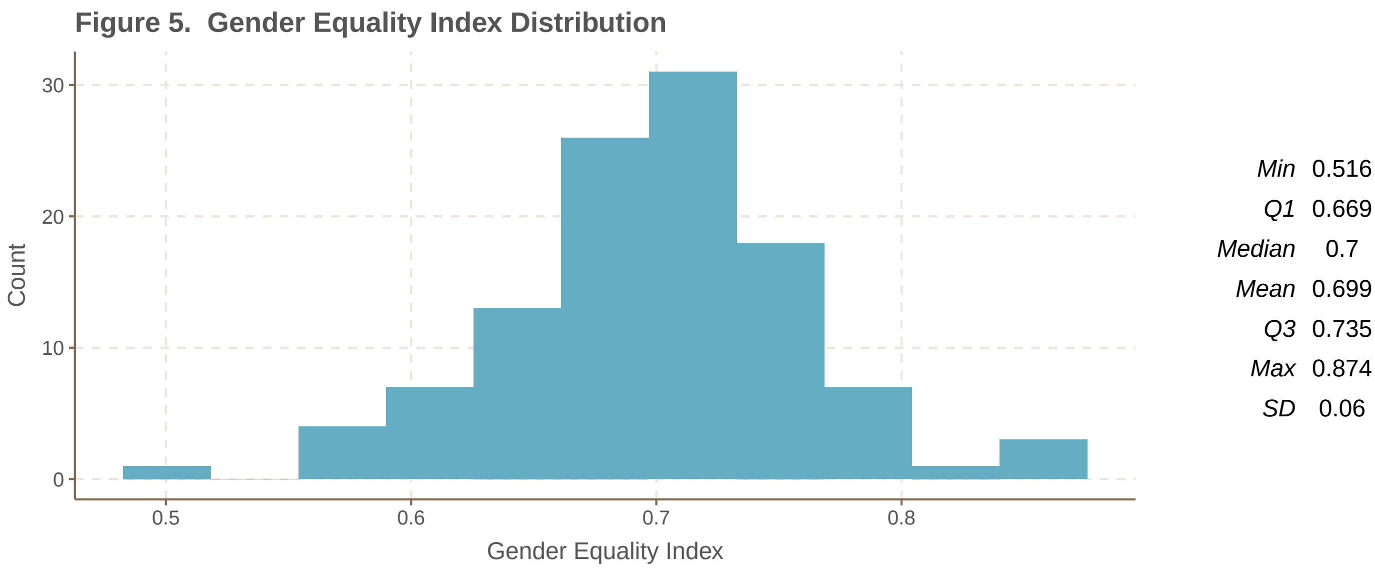


Figure 5 shows a roughly symmetric, possibly trimodal distribution for `gendereq`, with mean and median very close in value, and standard deviation approximately a tenth of either measure of location.



The probability of infant mortality is heavily right-skewed per Fig. 6.

Figure 6. Infant Mortality Distribution

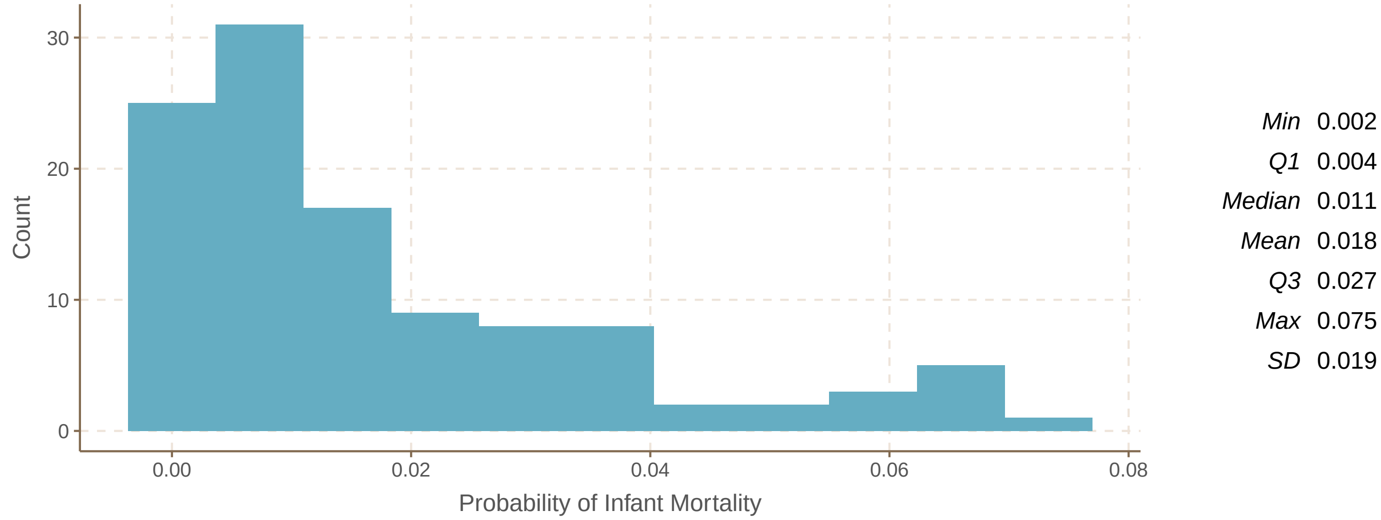
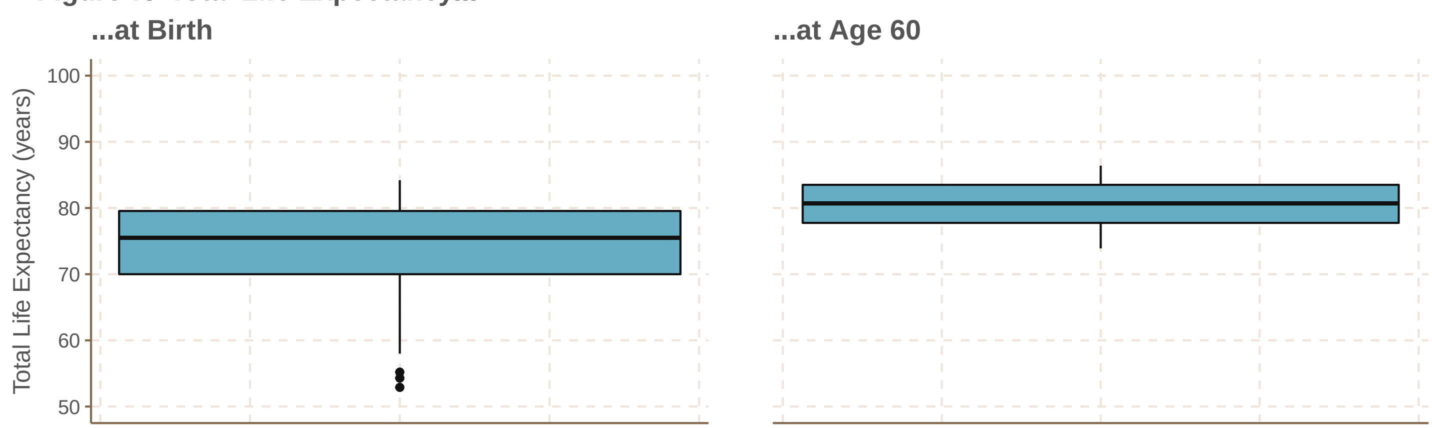


Figure 7 explores total life expectancy at birth vs. age 60. Life expectancy at birth appears left-skewed, but total life expectancy at sixty looks reasonably symmetric. These data indicate that, across all countries, a person who survives until at least 60 years has a longer total life expectancy than does the general world population at birth.

Figure 7. Total Life Expectancy...



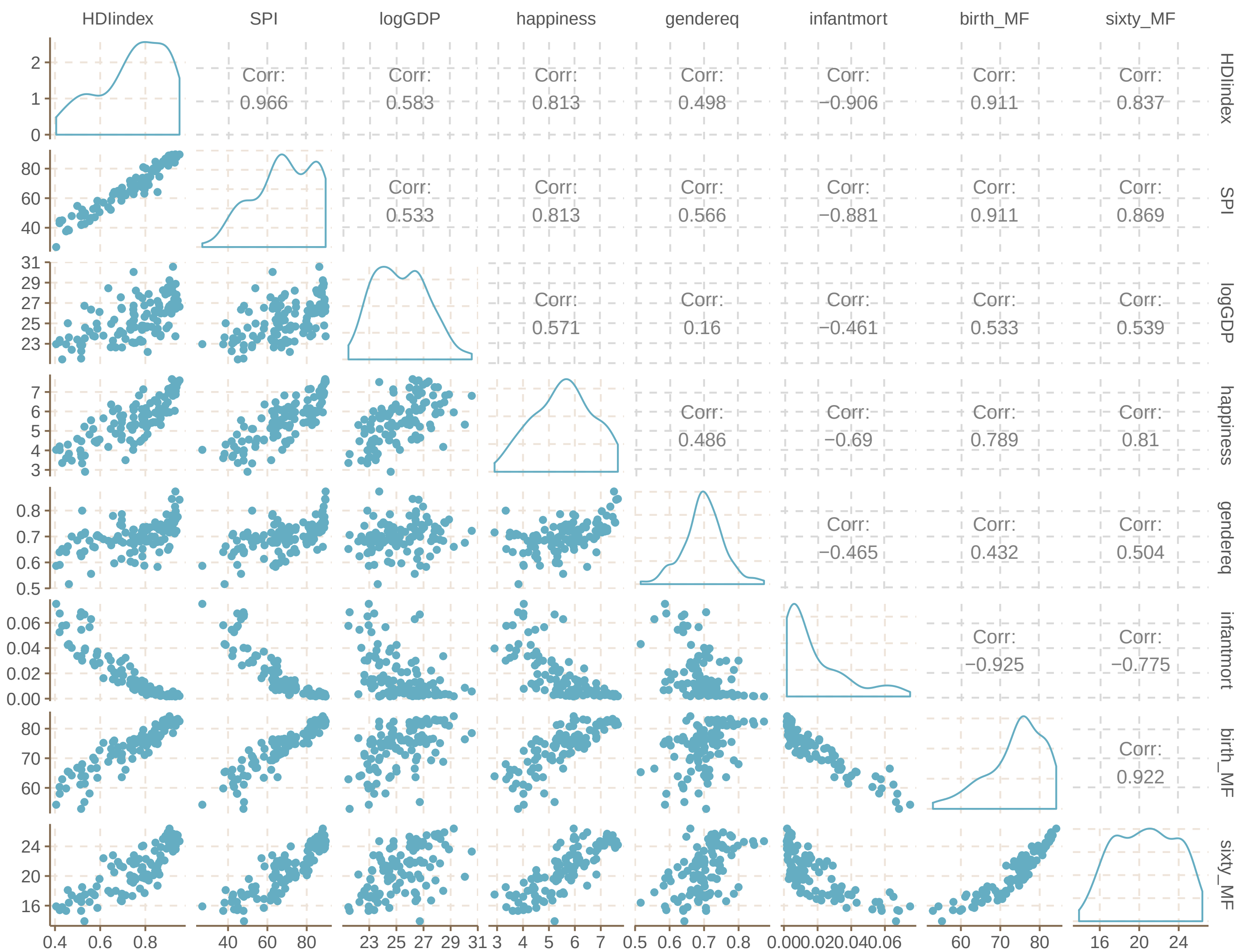
Min	Q1	Median	Mean	Q3	Max	SD
52.9	70	75.5	73.959	79.55	84.2	7.27

Min	Q1	Median	Mean	Q3	Max	SD
73.9	77.75	80.7	80.662	83.5	86.4	3.107

Figure 8 depcits pairwise relationships between continuous variables. Strong positive linear relationships are seen between `HDIindex` and `SPI`, `HDIindex` and `birth_MF`, and `SPI` and `birth_MF`, indicating that higher levels of social progress and human development generally correlate with a longer life expectancy at birth. Additionally, positive linear relationships are seen between `happiness` and `birth_MF`, between `happiness` and `HDIindex`, and between `happiness` and `SPI`, indicating that increasing levels of human development and social progress generally lead to increased happiness (though there is considerable spread in the data), and that happiness and longevity may also be correlated.

Strong negative relationships are seen between `infantmort` and `birth_MF`, between `HDIindex` and `infantmort`, and between `SPI` and `infantmort`, though the latter two of these may not necessarily be linear. These indicate that increasing human development and social progress levels correlate with lower levels of infant mortality, and that as infant mortality decreases, life expectancy at birth increases.

Figure 8. Correlation Matrix, Continuous Variables



Nonparametric and Parametric Analyses

For Analysis 1, the one-sample sign test yielded a p-value of $p_{sign} = 0.036$. Since $p_{sign} < \alpha$, the null hypothesis was rejected, leading to the conclusion that the median and mean HDI index values are not equal. However, the parametric test resulted in $p_t = 0.055$. Since $p_t > \alpha$, the null hypothesis was retained with the conclusion that, at the $\alpha = 0.05$ level, there is insufficient evidence to assert that the mean and median are different in the parametric case. The sign test is appropriate for this data; however, due to the clear nonnormality of the `HDIindex` data, the t-test is inappropriate because the assumption of normality is violated. Thus the nonparametric method results were retained.

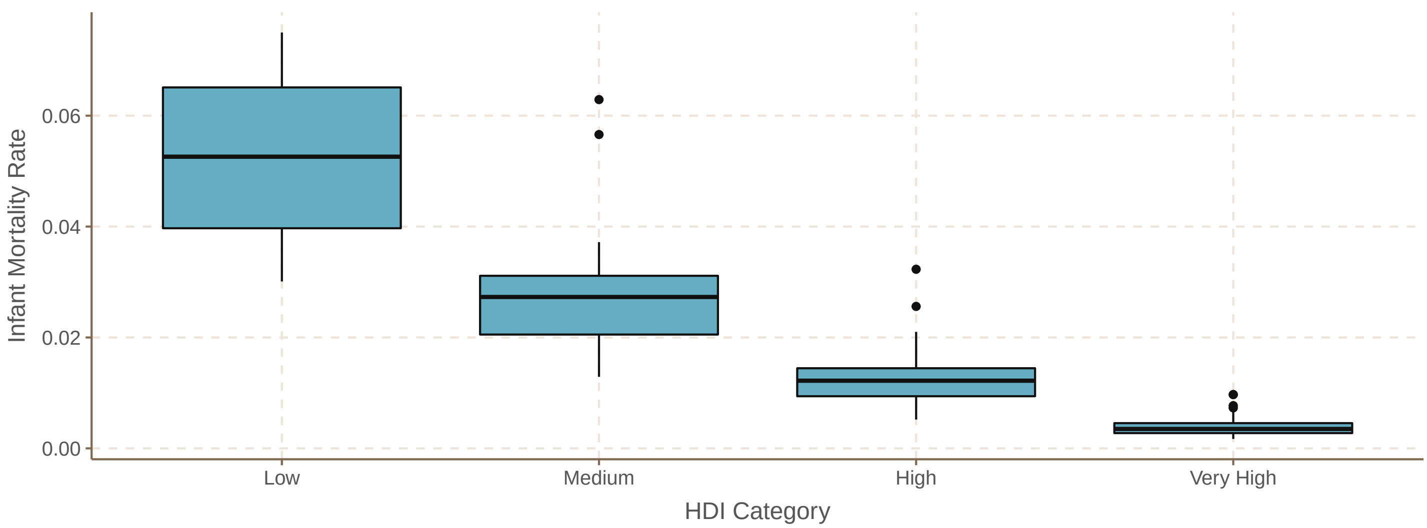
For Analysis 2, ties were not present in either `SPI` or `HDIindex`; thus, Kendall's Tau was tested using the usual `cor.test(x, y, method = "kendall")` approach. When testing for independence in both the nonparametric and parametrics cases, the p-values are less than 0.0001, with correlation coefficients $r_\tau = 0.843$ and $r = 0.966$ respectively. Both are statistically significant and represent strong evidence of a linear relationship between social progress and human developmental level, thus H_0 was rejected with the conclusion that, at the $\alpha = 0.05$ level, social progress and human development are correlated. Both tests assume continuous data, which is satisfied, and both yield the same p-value. The nonparametric method is more conservative, with a slightly smaller correlation coefficient for Kendall's Tau as compared to Pearson's, so the nonparametric method results were retained.

When testing for independence in Analysis 3 for both the nonparametric and parametrics cases, the p-values were less than 0.0001, leading to the rejection of H_0 (i.e., independence) and the conclusion that, at the $\alpha = 0.05$ level, there is sufficient evidence of an association between log(GDP) and infant mortality. The correlation matrix shows that this is an inverse relationship - i.e., as log(GDP) increases, infant mortality decreases. This relationship was sufficiently linear to yield a significant p-value on Pearson's test, so although the scatterplot in Fig. 8 appears nonlinear, it may in fact be linear. Either test is acceptable in this case.

In Analysis 4, the one-sample Kolmogorov-Smirnov test yielded p-value $p_{KS} = 0.915$, whereas the p-value for the Shapiro-Wilk test is $p_{SW} = 0.129$. Both exceed alpha; thus, H_0 is retained for both tests and I conclude that there is insufficient evidence to assert that `happiness` is not normally distributed. Both tests assume a continuous distribution, which is satisfied by the data, and the tests agree, so either the nonparametric or parametric approach is acceptable in this case. Notably, although neither p-value is significant, p_{SW} is much smaller than p_{KS}.

Both the permutation F-test and ANOVA yielded $p < 0.0001$ for Analysis 5. Since $p < \alpha$, H_0 was rejected for both the nonparametric and parametric cases, leading to the conclusion that there exists a difference in infant mortality by HDI category. The assumption of independence is satisfied by the data. However, per Fig. 9, the ANOVA assumption of constant variance is violated, and additionally, normality may be violated for the Medium and High categories. Therefore, the parametric method is inappropriate in this case, and the nonparametric method is retained.

Figure 9. Infant Mortality Rate by HDI Category



Discussion

Nonparametric methods are appropriate when analyzing data that violates the assumption of normality, and they can also be applied to normal data, but may be less powerful than the traditional parametric tests. The dataset used in this study consists of real-world QoL data, many variables of which fail normality. Across all five analyses, the nonparametric methods performed acceptably, whereas the parametric equivalents were only acceptable for two of the analyses.

THINK ABOUT WHAT ELSE TO ADD TO THE DISCUSSION - ADDRESS HOW UNDERSTANDING THE RELATIONSHIPS BETWEEN QOL MEASURES IS USEFUL

Limitations

An important limitation to this analysis is that countries with missing data were excluded from the analytic dataset. This may introduce bias, as a nontrivial proportion of countries omitted belong to the low and medium HDI developmental categories, and these countries are likely to perform poorly on the other QoL measures considered.

Another limitation is the inherent interrelatedness of some of these variables, particularly the compound indices. For example, one factor considered in the Human Development Index is life expectancy, so it is unsurprising that a strong association between `HDIindex` and `birth_MF` was seen in the correlation matrix. Similarly, some subitems of the Social Progress Index overlap with those included in the Gender Equality Index. Correlation tests are possible in such cases but may be inappropriate unless the underlying measures are fundamentally different.

Finally, this study is limited by the variables included in this dataset. There exist many other QoL measures that would have been worth including in this analysis, such as measures of educational attainment, civilian indebtedness, and population-level health utilities. Future work may entail expanding the study to include such variables.

Conclusion

Nonparametric methods represent an important toolset when working with data that violates normality. This analysis gives evidence that they can perform favorably as compared to their parametric equivalents, and should be considered when analyzing data that is unsuitable for parametric approaches.

References

Helliwell, John F., Richard Layard, and Jeffrey D. Sachs. 2018. "World Happiness Report." http://worldhappiness.report/ed/2018/.

Prioli, Katherine M. 2018a. "MAT_8452_Final_Project." https://github.com/kmprioliPROF/MAT_8452_Final_Project.

———. 2018b. "MAT_8790_Final_Project." https://github.com/kmprioliPROF/MAT_8790_Final_Project.

Social Progress Imperative. 2018a. "Social Progress Index." https://www.socialprogress.org/?tab=4.

———. 2018b. "Social Progress Index - Overview." https://www.socialprogress.org/index.

The United Nations Development Programme. 2018a. "Human Development Index." http://hdr.undp.org/en/data.

———. 2018b. "Human Development Reports - Human Development Index (HDI)." http://hdr.undp.org/en/content/human-development-index-hdi.

The World Bank. 2018. "Gross Domestic Product." https://data.worldbank.org/indicator/ny.gdp.mktp.cd?view=map&year_high_desc=true.

World Economic Forum. 2016a. "Gender Equality." http://reports.weforum.org/global-gender-gap-report-2016/rankings/.

———. 2016b. "Measuring the Global Gender Gap." http://reports.weforum.org/global-gender-gap-report-2016/measuring-the-global-gender-gap/.

World Health Organization. 2018a. "Life Expectancy." http://apps.who.int/gho/data/view.main.SDG2016LEXv?lang=en.

———. 2018b. "Probability of Dying Per 1000 Live Births." http://apps.who.int/gho/data/view.main.182?lang=en.