Mathematics Research Paper

TOPIC: To Determine an Association between Happiness and Human Development Index

Candidate Name : Anish Nitin Thakur
Session: May 2021

Introduction:

A few years back, I had travelled to a country called Bhutan. It is known to be one of the happiest and cleanest countries in the world, because it was probably the only carbon negative country at that time. So I kept wondering how it was possible, that a country with such limited resources could achieve such great standards. This question always remained inside my mind and made me curious. As the world is going through various types of global issues, I was amazed with this fact that a small country like Bhutan could achieve such a milestone, that even big developed countries could not. This curiosity made me want to learn about these parameters even more. I wanted to try to find out more about this because somewhere I thought that maybe development and happiness were aspects that could be co-related.

Aim: To investigate whether Happiness Index **(HPI)** and Human Development Index **(HDI)** have any co-relations (*Report Year: 2016)*

Objective and Plan of Action: To discover associations between The Happiness Index of a country and The Human Development Index of a country using mathematical methodologies, some of which include *Scatter Diagrams, Regression of Line Equations and The Chi Square Test*, to help me analyze the data and come up with a reliable conclusion.

Key Terms & Variables:

The Global Happiness Organization is a public organization that aims to support happiness and reduce suffering worldwide. It is presented in what is known as **The World Happiness Report**. This report is a yearly publication by the United Nation's Sustainable Development Solutions Network department. It comprises of articles, and ranks of national happiness established on ratings of the country's population by taking into consideration of various life factors. The findings refer to *six fundamental variables* to come to a reliable score. These include income, freedom, trust in government, healthy life expectancy, social support from family and friends, and generosity. Its main purpose is to examine and evaluate the science of computing and consider the subjective well-being, and track the quality of lives being lived in more than 150 countries. The **Happiness Index** stated in the report is simply calculated by taking an average score of all these factors to a single number score.

Human development is defined as the process of expanding people's liberties and opportunities and improving their welfare. Human development is about the real freedom for ordinary people. They have to decide who to be, what to do, and how to live. This concept was developed by an economist named *Mahbub ul Haq*. Human development is also characterized by the variation of material conditions.

The Human Development Index is a statistic composite index of life expectancy, education, and per capita income indicators, which are used to rank countries into four tiers of human development and all the others mentioned above.

The data I have collected is for over 130 countries regarding their Happiness Index and Human Development Index. This is a report annually published known as the World Happiness for the Year 2016. All my calculations are based on each data set collected for each of the countries which come under different regions around the world.

The Happiness Index is denoted as *x*, and the Human Development Index is denoted as *y* in the following data table.
The first piece of calculation I will be carrying out using these data sets, is to find the Co-Relation between the variables HDI (x) and HPI (y) in the given Raw Data Table below.

Raw Data Table						
	HPI (x value)	HDI (y value)	x^2	y^2	$x \times y$	REGION
Canada	23.9	0.922	571.21	0.850084	22.0358	Americas
United States of America	20.7	0.922	428.49	0.850084	19.0854	Americas
Chile	31.7	0.842	1004.89	0.708964	26.6914	Americas
Argentina	35.2	0.822	1239.04	0.675684	28.9344	Americas
Uruguay	36.1	0.802	1303.21	0.643204	28.9522	Americas
Costa Rica	44.7	0.792	1998.09	0.627264	35.4024	Americas
Panama	39.5	0.785	1560.25	0.616225	31.0075	Americas
Trinidad and Tobago	15.6	0.785	243.36	0.616225	12.246	Americas
Mexico	40.7	0.772	1656.49	0.595984	31.4204	Americas
Brazil	34.3	0.758	1176.49	0.574564	25.9994	Americas
Ecuador	37	0.749	1369	0.561001	27.713	Americas
Peru	34.6	0.748	1197.16	0.559504	25.8808	Americas
Colombia	40.5	0.746	1640.25	0.556516	30.213	Americas
Dominican Republic	30.3	0.733	918.09	0.537289	22.2099	Americas
Jamaica	36.9	0.732	1361.61	0.535824	27.0108	Americas
Suriname	25.4	0.719	645.16	0.516961	18.2626	Americas
Belize	33.8	0.708	1142.44	0.501264	23.9304	Americas
Paraguay	23.3	0.702	542.89	0.492804	16.3566	Americas
Bolivia	23.4	0.689	547.56	0.474721	16.1226	Americas
El Salvador	35.6	0.679	1267.36	0.461041	24.1724	Americas
Nicaragua	38.7	0.657	1497.69	0.431649	25.4259	Americas
Guatemala	34.3	0.649	1176.49	0.421201	22.2607	Americas
Honduras	27.2	0.614	739.84	0.376996	16.7008	Americas
Haiti	28.6	0.496	817.96	0.246016	14.1856	Americas
Australia	21.2	0.938	449.44	0.879844	19.8856	Asia Pacific
Hong Kong	16.8	0.93	282.24	0.8649	15.624	Asia Pacific
New Zealand	31.3	0.915	979.69	0.837225	28.6395	Asia Pacific
Japan	28.3	0.907	800.89	0.822649	25.6681	Asia Pacific

South Korea	24.8	0.9	615.04	0.81	22.32	Asia Pacific
Malaysia	30.3	0.799	918.09	0.638401	24.2097	Asia Pacific
Sri Lanka	33.8	0.768	1142.44	0.589824	25.9584	Asia Pacific
China	25.7	0.748	660.49	0.559504	19.2236	Asia Pacific
Thailand	37.3	0.748	1391.29	0.559504	27.9004	Asia Pacific
Mongolia	14.3	0.743	204.49	0.552049	10.6249	Asia Pacific
Philippines	35	0.696	1225	0.484416	24.36	Asia Pacific
Indonesia	35.7	0.691	1274.49	0.477481	24.6687	Asia Pacific
Vietnam	40.3	0.689	1624.09	0.474721	27.7667	Asia Pacific
India	29.2	0.636	852.64	0.404496	18.5712	Asia Pacific
Bhutan	28.6	0.609	817.96	0.370881	17.4174	Asia Pacific
Vanuatu	40.6	0.6	1648.36	0.36	24.36	Asia Pacific
Bangladesh	38.4	0.597	1474.56	0.356409	22.9248	Asia Pacific
Cambodia	25.6	0.576	655.36	0.331776	14.7456	Asia Pacific
Myanmar	24.7	0.574	610.09	0.329476	14.1778	Asia Pacific
Nepal	30.5	0.569	930.25	0.323761	17.3545	Asia Pacific
Pakistan	31.5	0.56	992.25	0.3136	17.64	Asia Pacific
Norway	36.8	0.951	1354.24	0.904401	34.9968	Europe
Switzerland	34.3	0.943	1176.49	0.889249	32.3449	Europe
Germany	29.8	0.934	888.04	0.872356	27.8332	Europe
Ireland	30	0.934	900	0.872356	28.02	Europe
Iceland	31.1	0.933	967.21	0.870489	29.0163	Europe
n= 135						
Sum	**3569.8**	**97.028**	**102436.36**	**72.993564**	**2631.0594**	
Average	**26.56740741**	**0.718725926**	**758.7878519**	**0.540693067**	**19.48932889**	

NOTE:

**** The data collected is of great quantity, therefore only first 50 data sets were presented in this table. The later half of the data can be found in the Appendix.***

**** The total sum and average taken were for 135 data sets (a combination of data present here and in the appendix)***

Calculation of Co-Relation between the variables:

Data Variables:

- $n - Sample\ Size\ (135)$
- $x - Happiness\ Index$
- $y - Human\ Development\ Index$
- $\sum x -\ Sum\ of\ all\ Happiness\ Index\ Values$
- $\sum y -\ Sum\ of\ all\ Human\ Development\ Index\ Values$
- $\sum xy - Sum\ of\ (Happiness\ Index\ .\ Human\ Development\ Index) Values$
- $\sum x^2 -\ Sum\ of\ all\ squares\ for\ x\ values\ (HPI)$
- $\sum y^2 -\ Sum\ of\ all\ squares\ for\ y\ value\ (HDI)$

Formula to Find Co-efficient of Correlation (r):

$$r = \frac{n\Sigma xy - \Sigma x.\Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2} \times \sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

From the above Table of Raw Data, I have found out the following values:

$\boldsymbol{\Sigma x^2} -\ 102436.4$

$\boldsymbol{\Sigma y^2} -\ 72.99356$

$\boldsymbol{\Sigma x.\Sigma y} -\ 2631.0594$

Line of Regression:

Equation: $\boldsymbol{y = ax + b}$

A regression line is a line in statistics that best represents the action of a collected data.
By fitting a linear equation to observed data, linear regression attempts to model the relationship between any two variables being compared. One variable is regarded as an independent variable, while the other is regarded as a dependent variable

Sample Variables and Equations:

Formula to Find Coefficient of Regression a

-

$$a = \frac{n\Sigma xy - \Sigma x . \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

* ($\boldsymbol{a}$ is the gradient of the regression line)

The **slope** specifies the gradient of a **line** and the **intercept** reveals the location at which the gradient overlaps with the vertical axis. The **slope** and the **intercept** help outline the linear association between the two variables. This could also be used in approximation for average rate of change.

Formula to Find the Vertical Intercept $\boldsymbol{b}$**:**

$$b = \frac{\Sigma y . \Sigma x^2 - \Sigma x . \Sigma xy}{n\Sigma x^2 - (\Sigma x)^2}$$

* ($\boldsymbol{b}$ is also known as beta)

Sample Calculations:

Equation to Find (**a**):

$$a = \frac{n\Sigma xy - \Sigma x.\Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

$$\frac{135(2631.059) - (3586.6\,.\,97.028)}{135(102436.4) - (3586.6)^2} = 0.0075$$

$$\therefore\ \boldsymbol{a} = \mathbf{0.0075}$$

Equation to Find (**b**):

$$b = \frac{\Sigma y.\Sigma x^2 - \Sigma x.\Sigma xy}{n\Sigma x^2 - (\Sigma x)^2}$$

$$\frac{(97.028\,.\,102436.4) - (3586.6\,.\,2631.059)}{135(102436.4) - (3586.6)^2} = 0.5208$$

$$\therefore\ \boldsymbol{b} = \mathbf{0.5208}$$

Calculation of Correlation Co-efficient:

$$r = \frac{n\Sigma xy - \Sigma x.\Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}.\ \sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

$$\frac{135(2631.059) - (3586.6\,.\,97.028)}{\sqrt{135(102436.4)} - (3586.6)^2\,.\sqrt{}(135(72.99356) - (97.028)^2} = 0.349125171$$

$$\therefore\ \boldsymbol{r} = \mathbf{0.349125171}$$

$$R^2 = 0.349125171\ ^2$$

$$\therefore\ \boldsymbol{R^2} = \mathbf{0.012189}$$

As we know that the line of best fit passes through the mean values of both the variables, therefore we can also calculate the value of ***b*** by substituting the mean values of variables ***x*** and ***y*** and the value of ***a*** obtained from the formula mentioned above. (Working shown on the right hand side)

$$\boldsymbol{y = ax + b}$$

$$\bar{x} = 3569.8/135 = 26.442$$

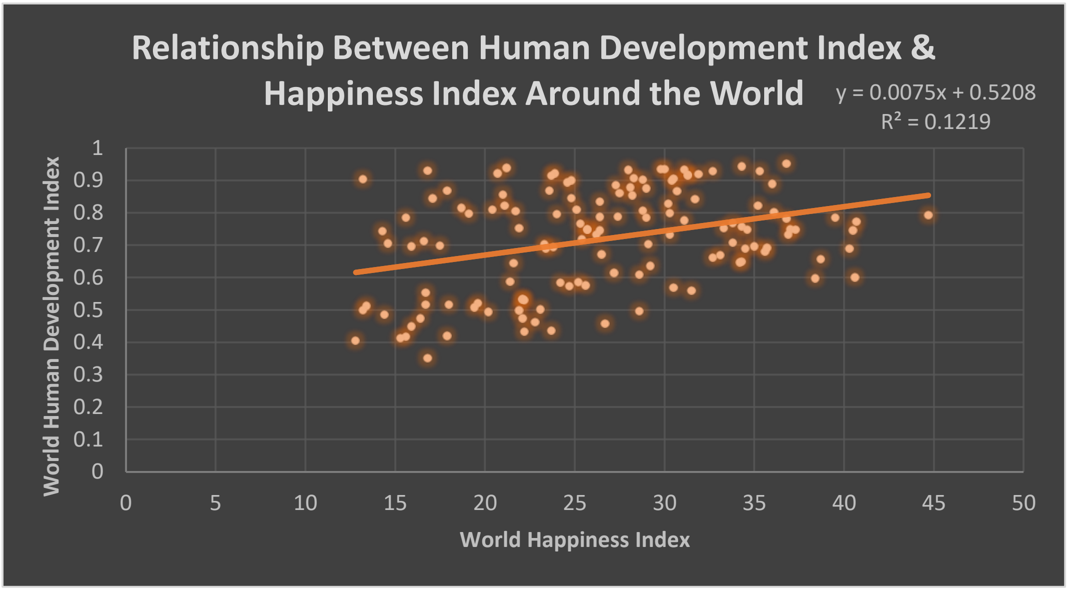
$$\bar{y} = 97.028/135 = 0.718$$

$$\mathbf{a = 0.0075}$$

$$\mathbf{0.718 = 0.0075 \times 26.442} + \boldsymbol{b}$$

$$\therefore\ \boldsymbol{b} = \mathbf{0.520}$$

Graphical Representation of Data (Line of Regression/ Best Fit)



R^2 **is** the square of the coefficient of correlation and indicates the percentage of variation. This value tends to increase as additional predictors in the model are included

The steps to calculate the R^2 value involve taking the dependent and independent variables' data points (observations) and evaluating the best match line, which is often extracted from a regression model as shown through. Then we must square the results by subtracting the true values from the expected values. The calculated values of the expected and observed results is shown below through the Observed and Expected Value Tables.

The "strong" or "good" R^2 value can vary depending on the situation. Even a low R-Squared, such as 0.1 to 0.5, can be found high in certain areas, such as the social sciences. In my case I have obtained an R^2 value of 0.12189. This could be considered as a strong correlation because a topic like this may not have a greater value due to the type of data that it is.

The next piece of calculation I will be carrying out is the <u>Chi Square Test</u>, to help me investigate the dependence or the independence of the variables used in my data.

Chi-Square Test for Independence:

A chi-square ($\chi 2$) statistic is a test, that examines the dissimilarity of a model with the help of factual observed data. Data used to measure a chi-square statistic must be random, raw, mutually exclusive, derived from independent variables and acquired from an adequately huge sample. In my case, the two variables of the data collected are raw and mutually exclusive, because the two indexes have different sets of parameters that are compared with to obtain a final value.

The Chi Squared value is also used in hypothesis testing. The chi-square data compares the scale of any differences between the **expected** outcomes and the **observed** outcomes. Provided the sample's size and the number of relationship variables, degrees of freedom are used for these investigations to decide whether, based on the total number of variables and samples within the investigation, a certain null hypothesis can be dismissed or not. The bigger the range of data, the more precise the outcomes would be, as with any statistical data generally.

In relation with my exploration, I purposely took a large sample size of data so that I could obtain a more accurate result and relation overall.

The formula of the *Chi-Squared Test for Independence* is:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

χ^2 = chi squared
O_i = observed value
E_i = expected value

There are two major types of chi-square tests:

- **the test of independence**, which poses a question of relationship in my case, such as, "Is there a relationship between happiness index and human development index?";
- **the goodness-of- fit test**, which asks something like, "To what extent are happiness and human development index actually related."

To make a contingency table, I will be classifying the data sets into three section. The first will contain data of the countries that have Happiness Index smaller than or equal to **(HPI ≤ 20).** The second section would be Happiness index of countries greater than 20 but less than 40 **(20 < HPI < 40**). And the last section will comprise of Happiness Index data of countries greater than 40 **(HPI > 40)**.

Independence:

An (χ2) **independence test** will inform us how likely the random chance will explain any observable discrepancy between the real data frequencies and these theoretical expectations. The null and alternative hypotheses are stated in words in a test of independence. Since there are two variables in the contingency table, the null hypothesis states that they are separate (independent), while the alternative hypothesis states that they are not (dependent). In relation to my findings, the hypotheses that can be formed are:

H_0 **(Null Hypothesis)**: "[Happiness Index] is independent of [Human Development Index]".

H_1 **(Alternative Hypothesis):** "[Happiness Index] is not independent of [Human Development Index]".

Hypothesis Testing:

The degrees of freedom for the chi-square are calculated using the following formula:

$$df = (r-1)(c-1)$$

Where r is the number of rows and c is the number of columns. If the observed chi-square test statistic is greater than the critical value, then the null hypothesis can be rejected.

There are 6 rows (r) and 3 columns (c) present in my data set provided on page 9. Therefore, the degree of freedom of my data is:

$$df = (r-1)(c-1)$$

$$df = (6-1)(3-1) = 10$$

$$\therefore \boldsymbol{df} = \mathbf{10}$$

Goodness-of-Fit

The Chi-Square goodness of fit test is a non-parametric test to evaluate how substantially the observed value of a phenomenon varies from the predicted value. The word goodness of fit is used to compare the observed sample distribution with the predicted probability distribution in the Chi-Square goodness of fit test. In this case it would allow me to find that to what extent are happiness and human development index related?

Sample Calculation:

Expected Value for America with HPI ≤ 20

$$\frac{\boldsymbol{Total\ of\ all\ HPI\ for\ America\ \times\ Total\ HPI\ \leq 20\ for\ all\ countries}}{\boldsymbol{Grand\ Total}}$$

$$\frac{24\ \times\ 27}{135} = 4.80\ (Expected\ Value)$$

<u>Observed Value Table</u>

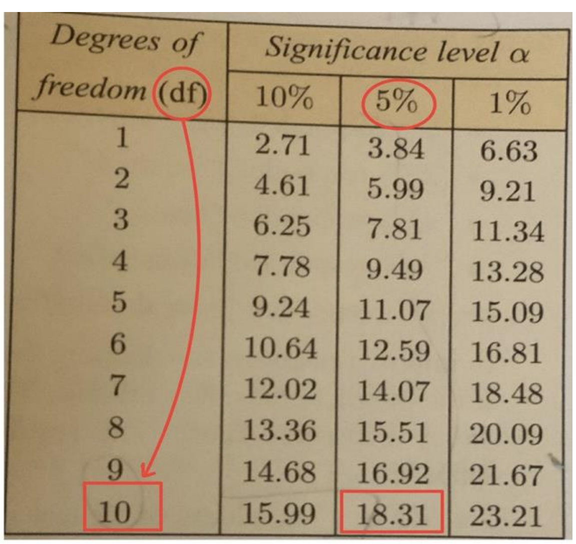
Region	HPI ≤ 20 x ≤ 20	20 < HPI < 40 20 < x <40	HPI > 40 x > 40	Total
America	1	20	3	24
Asia Pacific	2	17	2	21
Europe	1	19	0	20
Middle East & North Africa	0	13	0	13
Post Communist	5	20	0	25
Sub Saharan Africa	18	14	0	32
Total	27	103	5	135

<u>Expected Value Table</u>

Region	HPI ≤ 20 x ≤ 20	20 < HPI < 40 20 < x < 40	HPI > 40 x > 40	Total
America	4.80	18.31	0.35	23.46
Asia Pacific	4.20	16.02	0.77	20.99
Europe	4.00	15.25	0.74	19.99
Middle East & North Africa	2.60	9.91	0.48	12.99
Post Communist	5.00	19.07	0.92	24.99
Sub Saharan Africa	6.40	24.41	1.18	31.99
Total	27	102.97	4.44	134.41

<u>Chi Square Test Table</u>

Region	**Observed Value [OB]**	**Expected Value [EX]**	**$\frac{(OB-EX)^2}{EX}$**
HPI ≤ 20	**HPI ≤ 20**	**HPI ≤ 20**	**HPI ≤ 20**
America	1	4.8	3.008333
Asia Pacific	2	4.2	1.152381
Europe	1	4.0	2.25
Middle East & North Africa	0	2.6	2.60
Post-Communist	5	5	0
Sub Saharan Africa	18	6.4	21.025
20 < HPI < 40	**20 < HPI < 40**	**20 < HPI < 40**	**20 < HPI < 40**
America	20	18.31	0.155986
Asia Pacific	17	16.02	0.05995
Europe	19	15.25	0.922131
Middle East & North Africa	13	9.91	0.963481
Post-Communist	20	19.07	0.045354
Sub Saharan Africa	14	24.41	4.439496
HPI > 40	**HPI > 40**	**HPI > 40**	**HPI > 40**
America	3	0.888	5.023135
Asia Pacific	2	0.77	1.964805
Europe	0	0.74	0.74
Middle East & North Africa	0	0.48	0.48
Post-Communist	0	0.92	0.92
Sub Saharan Africa	0	1.18	1.18
Sum	$\therefore\ \boldsymbol{CHI\ SQUARE\ VALUE}(\chi 2)\ =\ \mathbf{46.93005}$		**46.93005**

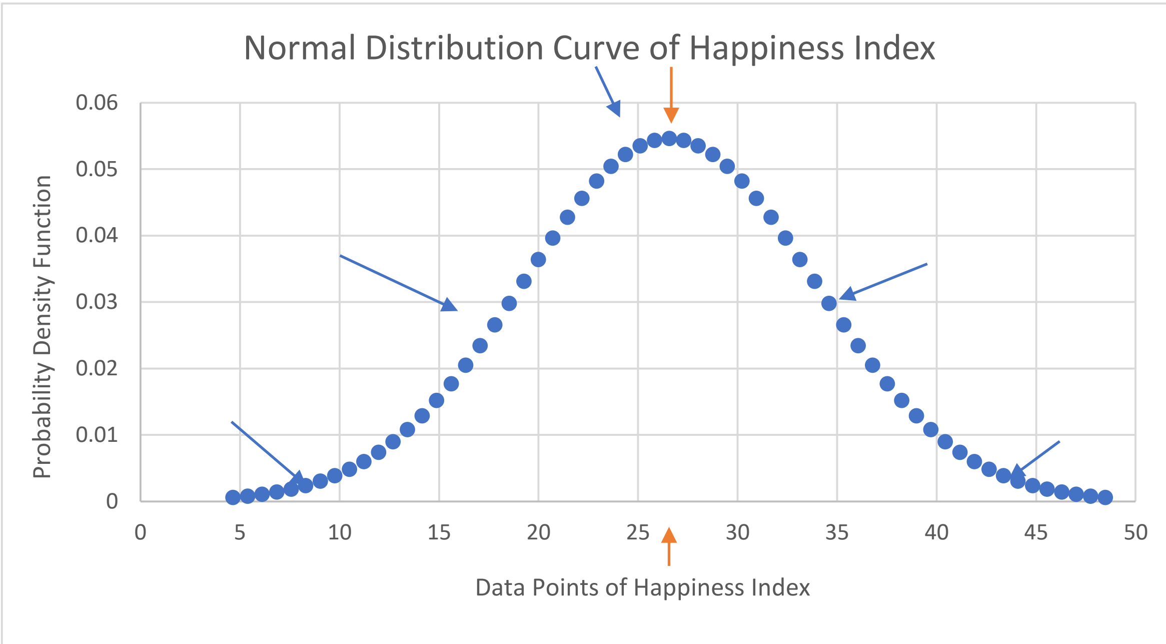


Degrees of freedom (df)	Significance level α 10%	5%	1%
1	2.71	3.84	6.63
2	4.61	5.99	9.21
3	6.25	7.81	11.34
4	7.78	9.49	13.28
5	9.24	11.07	15.09
6	10.64	12.59	16.81
7	12.02	14.07	18.48
8	13.36	15.51	20.09
9	14.68	16.92	21.67
10	15.99	18.31	23.21

Due to the sample size being quite limited, 95% is mostly considered as a strongly based evidence and therefore this is why I chose 0.05 or 5% as the Significance level. The degree of freedom obtained for the data set is 10. When this degree of freedom value is referred to left with significance level of 5% as shown in the table on the left side, the critical value shows 18.31.

This indicates that the critical value is < calculated value, thus allowing us to reject the null hypothesis. This states that the variables in the data table are not independent, which eventually means that the two variables are in fact related.

Following this, I wanted to see if my grouped data for happiness Index followed a normal distribution. I estimated the expected values using Excel. I found that some of my expected values were less than the observed ones. After graphing it, I found out that the data obtained for happiness index could be distributed normally as well. This shows that I could use this data set with another Mathematical Methodology, which thus indicates its versatility.



Sample Calculations for the Normal Distribution Curve, using Excel Sheet and Functions

		Mean	Std Dev		Z score	Data Point	Normalize
*	23.9	26.567407	7.3045154		Increments		
*	20.7				-3	4.6538613	0.0006067
*	31.7				-2.9	5.3843128	0.0008149
	35.2				-2.8	6.1147644	0.0010836
	36.1				-2.7	6.8452159	0.0014266
	44.7				-2.6	7.5756675	0.0018595
	39.5				-2.5	8.306119	0.0023997
	15.6				-2.4	9.0365705	0.0030658
	40.7				-2.3	9.7670221	0.003878
	34.3				-2.2	10.497474	0.0048565
	37				-2.1	11.227925	0.0060214
	34.6				-2	11.958377	0.0073915
	40.5				-1.9	12.688828	0.0089829
*	30.3				-1.8	13.41928	0.0108084
	36.9				-1.7	14.149731	0.0128755
*	25.4				-1.6	14.880183	0.0151852
*	33.8				-1.5	15.610634	0.0177312
*	23.3				-1.4	16.341086	0.0204979
*	23.4				-1.3	17.071537	0.0234606
	35.6				-1.2	17.801989	0.0265844
	38.7				-1.1	18.532441	0.0298243
	34.3				-1	19.262892	0.0331262
*	27.2				-0.9	19.993344	0.0364275
*	28.6				-0.8	20.723795	0.0396592
*	21.2				-0.7	21.454247	0.0427481
	16.8				-0.6	22.184698	0.045619
*	31.3				-0.5	22.91515	0.0481983
*	28.3				-0.4	23.645601	0.0504168
*	24.8				-0.3	24.376053	0.0522126
*	30.3				-0.2	25.106504	0.0535344
*	33.8				-0.1	25.836956	0.0543434
*	25.7				0	26.567407	0.0546158

	37.3	0.1	27.297859	0.0543434
	14.3	0.2	28.02831	0.0535344
	35	0.3	28.758762	0.0522126
	35.7	0.4	29.489214	0.0504168
	40.3	0.5	30.219665	0.0481983
*	29.2	0.6	30.950117	0.045619
*	28.6	0.7	31.680568	0.0427481
	40.6	0.8	32.41102	0.0396592
	38.4	0.9	33.141471	0.0364275
*	25.6	1.0	33.871923	0.0331262
*	24.7	1.1	34.602374	0.0298243
*	30.5	1.2	35.332826	0.0265844
*	31.5	1.3	36.063277	0.0234606
	36.8	1.4	36.793729	0.0204979
	34.3	1.5	37.52418	0.0177312
*	29.8	1.6	38.254632	0.0151852
*	30	1.7	38.985084	0.0128755
*	31.1	1.8	39.715535	0.0108084
*	28	1.9	40.445987	0.0089829
*	32.7	2.0	41.176438	0.0073915
	35.3	2.1	41.90689	0.0060214
*	31.9	2.2	42.637341	0.0048565
*	31.3	2.3	43.367793	0.003878
*	23.7	2.4	44.098244	0.0030658
*	30.5	2.5	44.828696	0.0023997
	13.2	2.6	45.559147	0.0018595
*	30.4	2.7	46.289599	0.0014266
	36	2.8	47.02005	0.0010836
*	28.1	2.9	47.750502	0.0008149
*	29	3.0	48.480954	0.0006067

** The points marked by an arrow on the Normal Distribution Curve represents the highlighted values as shown in the sample working above.*

Normal Distribution Graph Calculations:

X- *Happiness Index*, **μ**- *Mean of the data sets*, **σ-** *Standard Deviation*

$X \sim N(\mu,\sigma^2)$

$X \sim N(26.567, 7.304^2)$

$\mu-\sigma = 26.567-7.304$
$= 19.263$

$\mu+\sigma = 26.567+7.304$
$= 33.871$

**Percentage of data points of population that come under the ranges (μ-σ) and (μ+σ):*

$$\frac{84}{135} \times 100 = 62\%$$

I also wanted to check the percentage of data for the Happiness Index that come under the range of ($\mu-2\sigma$ to $\mu+2\sigma$).

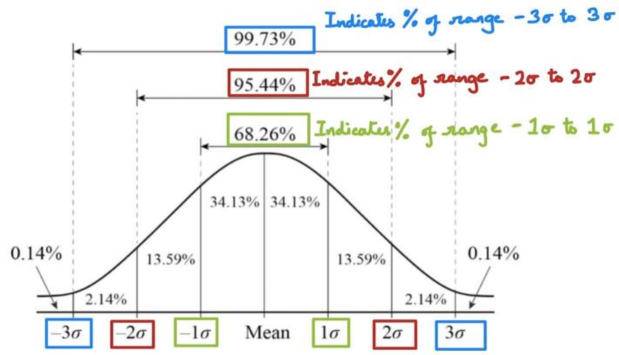
$\mu-2\sigma = 26.567 - 2(7.304)$
$= 11.959$

$\mu+2\sigma = 26.567 + 2(7.304)$
$= 41.175$

*Percentage of data points of population that come under the ranges *(μ-2σ) and (μ+2σ):*

$$\frac{134}{135} \times 100 = 99.25\%$$

The figure below shows the percentages of the accepted or literature values for all the three variation for the ranges of the mean and the standard deviation. This diagram of the Normal Distribution Curve was taken from a Mathematics Textbook and was then annotated by me.



Compared to the literature figure the data of my findings is limited. The percentage that I am getting is 62.0%, whereas the standard value should be around 68.4%, for the range (μ-σ and μ+σ). A reason for this slight difference may be due to the fact that my data is of limited number. This just shows that most of the countries that come under the Happiness Index follow the normal distribution.

The 99.25% found in the range (μ-2σ and μ+2σ) is quite a high percentage, whereas the accepted value should be around 95.44% . The reason behind this may be because I only took data for a set of 135 countries, whereas there are more than 200 countries around the globe. This may be taken as one of the limitations of my investigation, thus resulting in varied results compared to the literature values.

Conclusion:

Overall the investigation to discover associations between The Happiness Index of a country and The Human Development Index of a country using mathematical methodologies was successful. The *Scatter Diagram* helped to provide a visual representation of the data collected, which in a way made it easier to make out at a first glance of the fact that if there was at all any relationship between the data collected for the two variables. Next, the *Regression of Line Equations* helped me come up with specific methods and calculations to prove whether the visually represented data/figure had a Linear result to it, and whether it showed a weak or strong correlation, which thus tells us how strong the relationship between the two variables are. Lastly *The Chi Square Test* came with a lot of calculations which helped make sure in finding out the extensity of the relationship between the two variables and also enabled me to reject the null hypothesis. This states that the variables in the data table are not independent, which eventually means that the two variables were in fact related. The Normal Distribution Curve was another method used to make sure my findings were correct and justified. The fact that I could use all these methods to prove the relationship between the two parameters chosen, shows the versatility of the data obtained.

I feel that the overall way I approached this topic and documented my work is clear enough to support my hypothesis. Although most of my findings and calculations seem to be correct, there was one modification I had to make at the starting stages of this report, which is something that would have affected my whole report, if not rectified. Initially I had taken the data sets for 100 countries from the World Report for the year 2018. After carrying out the calculations on the data I had collected, I found out that I was not receiving the results that I needed to get, as the calculated values differed from the expected values and all the Regression Line Equations. Therefore, after carrying out some more detailed research, I came across the Report for the Year 2016. After carrying out the same pieces of calculations on this data set as mentioned thoroughly in this report, I finally got the results I had expected. This is the reason why I decided to base my investigation on the World Report for the year 2016. Another limitation to my investigation would be the fact that I only took data for a set of 135 countries, whereas there are more than 200 countries around the globe, thus resulting in varied results compared to the literature values.

The relationship between the Happiness Index and the Human Development Index can be used by different sectors of the government of a country. Countries can use it to help them reflect on their previous performances and also help to bring about changes in the different sections and parameters that are taken account into calculating their overall score for both the Happiness Index and the Human Development Index. This would be especially helpful in raising the standard of living in the country which thus will contribute to an overall development and economic improvement in a country.

Extension:

A possible extension for this investigation would be to firstly find out other parameters that define a country's overall growth. Next, I can take other two parameters and put them as my x and y variables, so that they could be compared. I can then carry out the same sets of calculations as shown in this report, that will help me find out whether the two parameters are related, and if they are then to what extent are they interconnected.

* Some examples of these parameters would be Natural Resources, Physical Capital or Infrastructure, Human Capital, Technology and Law.

Bibliography and References:

1. *World Happiness Report 2016*, worldhappiness.report/ed/2016/.
2. Jan-Emmanuel De Neve University of OxfordChristian Krekel London School oEconomics, et al. *Cities and Happiness: A Global Ranking and Analysis*, 20 Mar. 2020, worldhappiness.report/ed/2020/cities-and-happiness-a-global-ranking-and-analysis/.
3. "Human Development Reports." *Human Development Index (HDI) | Human Development Reports*, hdr.undp.org/en/content/human-development-index-hdi.
4. "Chi-Square Statistic: How to Calculate It / Distribution." *Statistics How To*, 10 Dec. 2020, www.statisticshowto.com/probability-and-statistics/chi-square/.
5. Barbara Illowsky & OpenStax et al. "Introduction to Statistics." *Lumen*, courses.lumenlearning.com/introstats1/chapter/test-of-independence/.
6. Stephanie. "Goodness of Fit Test: What Is It?" *Statistics How To*, 16 Sept. 2020, www.statisticshowto.com/goodness-of-fit-test/.
7. Hease, Michael, et al. *Mathematics: Applications and Interpretation SL*. Hease Mathematics, 2019.
8. Bansal, -- By Sumit, et al. "How to Make a Bell Curve in Excel (Step-by-Step Guide)." *Trump Excel*, 30 Nov. 2020, trumpexcel.com/bell-curve/.
9. *Google Search*, Google, www.google.com/search?q=normal%2Bdistribution%2Bpercentages%2Bdiagram&source=lnms&tbm=isch&sa=X&ved=2ahUKEwiTv8SG1u7vAhVRbSsKHbd8C_sQ_AUoAXoECAEQAw&biw=1085&bih=745#imgrc=oNjC01lnaMFSrM.

Appendix

Data Set for the remaining 85 countries

Sweden	28	0.932	784	0.868624	26.096	Europe
Denmark	32.7	0.928	1069.29	0.861184	30.3456	Europe
Netherlands	35.3	0.928	1246.09	0.861184	32.7584	Europe
United Kingdom	31.9	0.92	1017.61	0.8464	29.348	Europe
Finland	31.3	0.918	979.69	0.842724	28.7334	Europe
Belgium	23.7	0.915	561.69	0.837225	21.6855	Europe
Austria	30.5	0.906	930.25	0.820836	27.633	Europe
Luxembourg	13.2	0.904	174.24	0.817216	11.9328	Europe
France	30.4	0.899	924.16	0.808201	27.3296	Europe
Spain	36	0.889	1296	0.790321	32.004	Europe
Italy	28.1	0.878	789.61	0.770884	24.6718	Europe
Malta	29	0.875	841	0.765625	25.375	Europe
Greece	23.6	0.868	556.96	0.753424	20.4848	Europe
Cyprus	30.7	0.867	942.49	0.751689	26.6169	Europe
Portugal	24.8	0.845	615.04	0.714025	20.956	Europe
Israel	28.8	0.902	829.44	0.813604	25.9776	ME & NA
Oman	21.1	0.822	445.21	0.675684	17.3442	ME & NA
Iran	24	0.796	576	0.633616	19.104	ME & NA
Turkey	26.4	0.787	696.96	0.619369	20.7768	ME & NA
Algeria	33.3	0.753	1108.89	0.567009	25.0749	ME & NA
Lebanon	21.9	0.753	479.61	0.567009	16.4907	ME & NA
Tunisia	26.2	0.732	686.44	0.535824	19.1784	ME & NA
Egypt	23.8	0.694	566.44	0.481636	16.5172	ME & NA
Palestine	34.5	0.689	1190.25	0.474721	23.7705	ME & NA
Iraq	26.5	0.672	702.25	0.451584	17.808	ME & NA
Morocco	32.7	0.662	1069.29	0.438244	21.6474	ME & NA
Afghanistan	20.2	0.494	408.04	0.244036	9.9788	ME & NA
Yemen	22.8	0.462	519.84	0.213444	10.5336	ME & NA
Slovenia	24.6	0.894	605.16	0.799236	21.9924	P-C
Czech Republic	27.3	0.885	745.29	0.783225	24.1605	P-C
Estonia	17.9	0.868	320.41	0.753424	15.5372	P-C
Poland	27.5	0.86	756.25	0.7396	23.65	P-C
Lithuania	21	0.855	441	0.731025	17.955	P-C
Slovakia	28.2	0.853	795.24	0.727609	24.0546	P-C
Latvia	17.1	0.844	292.41	0.712336	14.4324	P-C
Hungary	26.4	0.835	696.96	0.697225	22.044	P-C
Croatia	30.2	0.828	912.04	0.685584	25.0056	P-C
Russia	18.7	0.815	349.69	0.664225	15.2405	P-C
Bulgaria	20.4	0.81	416.16	0.6561	16.524	P-C
Montenegro	25.1	0.81	630.01	0.6561	20.331	P-C
Romania	28.8	0.807	829.44	0.651249	23.2416	P-C

Belarus	21.7	0.805	470.89	0.648025	17.4685	P-C
Kazakhstan	19.1	0.797	364.81	0.635209	15.2227	P-C
Serbia	29	0.785	841	0.616225	22.765	P-C
Albania	36.8	0.782	1354.24	0.611524	28.7776	P-C
Georgia	31.1	0.776	967.21	0.602176	24.1336	P-C
Bosnia and Herzegovina	25.3	0.766	640.09	0.586756	19.3798	P-C
Armenia	25.7	0.749	660.49	0.561001	19.2493	P-C
Ukraine	26.4	0.746	696.96	0.556516	19.6944	P-C
Turkmenistan	14.6	0.705	213.16	0.497025	10.293	P-C
Uzbekistan	29.1	0.703	846.81	0.494209	20.4573	P-C
Kyrgyzstan	33.1	0.669	1095.61	0.447561	22.1439	P-C
Tajikistan	34.2	0.647	1169.64	0.418609	22.1274	P-C
Mauritius	27.4	0.788	750.76	0.620944	21.5912	S-S-A
Botswana	16.6	0.712	275.56	0.506944	11.8192	S-S-A
Gabon	17.5	0.698	306.25	0.487204	12.215	S-S-A
South Africa	15.9	0.696	252.81	0.484416	11.0664	S-S-A
Namibia	21.6	0.645	466.56	0.416025	13.932	S-S-A
Ghana	21.4	0.588	457.96	0.345744	12.5832	S-S-A
Zambia	25.2	0.586	635.04	0.343396	14.7672	S-S-A
Kenya	24.2	0.585	585.64	0.342225	14.157	S-S-A
Cameroon	16.7	0.553	278.89	0.305809	9.2351	S-S-A
Tanzania	22.1	0.533	488.41	0.284089	11.7793	S-S-A
Zimbabwe	22.1	0.532	488.41	0.283024	11.7572	S-S-A
Nigeria	22.2	0.53	492.84	0.2809	11.766	S-S-A
Rwanda	19.6	0.52	384.16	0.2704	10.192	S-S-A
Lesotho	16.7	0.516	278.89	0.266256	8.6172	S-S-A
Mauritania	18	0.516	324	0.266256	9.288	S-S-A
Benin	13.4	0.512	179.56	0.262144	6.8608	S-S-A
Uganda	19.4	0.508	376.36	0.258064	9.8552	S-S-A
Comoros	23.1	0.502	533.61	0.252004	11.5962	S-S-A
Togo	13.2	0.5	174.24	0.25	6.6	S-S-A
Senegal	21.9	0.499	479.61	0.249001	10.9281	S-S-A
Cote d'Ivoire	14.4	0.486	207.36	0.236196	6.9984	S-S-A
Djibouti	16.4	0.474	268.96	0.224676	7.7736	S-S-A
Malawi	22.1	0.474	488.41	0.224676	10.4754	S-S-A
Ethiopia	26.7	0.457	712.89	0.208849	12.2019	S-S-A
Guinea	15.9	0.449	252.81	0.201601	7.1391	S-S-A
Mozambique	23.7	0.435	561.69	0.189225	10.3095	S-S-A
Liberia	22.2	0.432	492.84	0.186624	9.5904	S-S-A
Burkina Faso	17.9	0.42	320.41	0.1764	7.518	S-S-A
Burundi	15.6	0.418	243.36	0.174724	6.5208	S-S-A
Sierra Leone	15.3	0.413	234.09	0.170569	6.3189	S-S-A
Chad	12.8	0.405	163.84	0.164025	5.184	S-S-A
Niger	16.8	0.351	282.24	0.123201	5.8968	S-S-A