

Title: An analysis on the relationship between life expectancy and years of schooling

Submitted by: OKUBO, Christine Joy C.

LINEAR REGRESSION

DISCUSSION

A dataset on the average life expectancy, adult mortality, infant deaths, ..., years of schooling per country from 2000-2015 was compiled from both the World Health Organization (WHO) and the United Nations (UN) by Rajarshi (2015). For this project, the only data columns in concern are the life expectancy and years of schooling per country. I have also narrowed the scope of my study to reflect only the statistics from 2015 (most recent) since using data from multiple years would yield numerous data points and consequently, require a multiple regression analysis. Lastly, the larger dataset contains data for 193 countries but this project will only use the 173 countries ($n = 173$) that have listed numbers for both life expectancy and years of schooling.

This project aims to analyze the relationship between two quantitative variables. The independent variable (x) is the number of years of schooling. The dependent variable (y) is the life expectancy in years.

This project investigated the correlation between life expectancy and years of schooling based on the data from 173 countries (as shown in Appendix A) using the linear correlation coefficient, r . To obtain r , the covariance and standard deviations of both x and y were used. The calculated $r = 0.8182$, indicates high positive correlation and suggests strong dependency and usability for prediction between the two variables (Jaadi, 2019).

This project was also able to determine a regression line for the data which are useful for making predictions on life expectancy based on years of schooling. I used the method of least squares to obtain the line of best fit: $y = 2.2287x + 42.9016$. The slope tells us that for every increase of one year in schooling, we can expect an increase of 2.2287 years in life expectancy. The y-intercept suggests that when the years of schooling is zero, the life expectancy will only be 42.9016 years.

IMPORTANT VALUES

[Calculated using Excel from values in Appendix A]

- $\Sigma x = 2236.40$, $\Sigma y = 12406.30$
- $\Sigma x^2 = 30369.20$, $\Sigma y^2 = 900515.23$
- $\Sigma xy = 163629.77$
- $SS(x) = \Sigma x^2 - \frac{(\Sigma x)^2}{n} = 30369.20 - \frac{(2236.40)^2}{173} = 1458.8823$
- $SS(y) = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = 900515.23 - \frac{(12406.30)^2}{173} = 10825.7520$
- $SS(xy) = \Sigma xy - \frac{\Sigma x * \Sigma y}{n} = 163629.77 - \frac{2236.40 * 12406.30}{173} = 3251.4502$

LINEAR CORRELATION COEFFICIENT, r

Formula for r: $r = \frac{covar(x,y)}{sx * sy} = \frac{18.9038}{2.9124 * 7.9335} = 0.8182$

- $\Sigma(x - \bar{x})(y - \bar{y}) = 3251.4502$
- $covar(x, y) = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{n-1} = \frac{3251.4502}{173-1} = 18.9038$

Alternative formula for r: $r = \frac{SS(xy)}{\sqrt{SS(x) * SS(y)}} = \frac{3251.4502}{\sqrt{1458.8823 * 10825.7520}} = 0.8182$

Discussion:

The calculated linear correlation coefficient, $r = 0.8182$, indicates strong positive correlation between the average life expectancy (y) and years of schooling (x) per country. This suggests that the underlying relationship between the two variables under consideration is almost positively linear ($r = +1$). Additionally, the significance of the linear correlation coefficient tells us that the two variables are somewhat dependent and useful for making predictions about the other.

Since the two variables have a strong positive linear relationship, we expect that as one variable increases, the other variable also increases; as the years of schooling increases, it is likely that the average life expectancy for a country in the same population as our sample will also increase. Other lurking variables that may be influencing or emerge from the association between years of schooling and life expectancy include the GDP and economic output, lifestyle, and number of doctors per country in the population.

THE REGRESSION LINE

$$[\text{Manual}] \ y = 2.2287x + 42.9016$$

Using the Method of least squares,

The line of best fit: $\hat{y} = b_0 + b_1x$

The slope: $b_1 = \frac{SS(xy)}{SS(x)} = \frac{3251.4502}{1458.8823} = 2.2287$

The y-intercept: $b_0 = \frac{\Sigma y - b_1 \Sigma x}{n} = \frac{12406.30 - (2.2287 * 2236.40)}{173} = 42.9016$

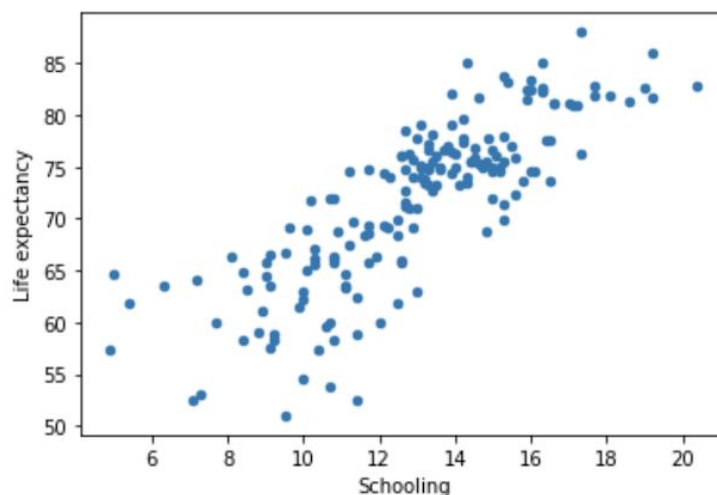
The regression line/line of best fit for life expectancy (y) vs. years of schooling (x) based from the data shown in Appendix A is $y = 2.2287x + 42.9016$. The slope of the line ($b_1 = 2.2287$) tells us that for every increase of one year in years of schooling, we can expect an increase of 2.2287 years in life expectancy in a country that is part of the population of concern. The y-intercept ($b_0 = 42.9016$) suggests that when the years of schooling is down to zero, we can only expect an average of 42.9016 years of life.

$$[\text{Software}] \ y = 2.22873x + 42.90159$$

Figure 1

Life Expectancy vs. Schooling across 173 countries in 2015

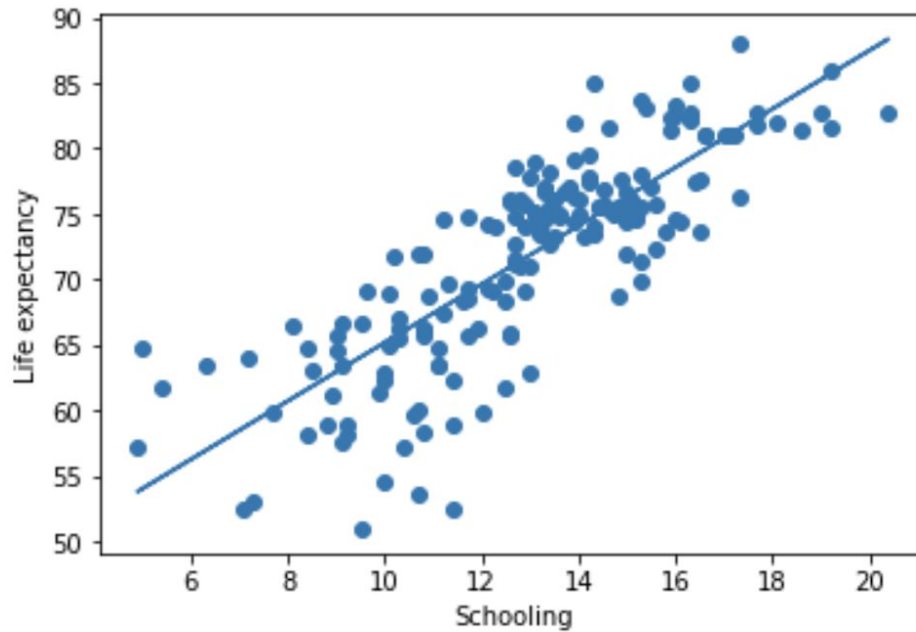
```
In [6]: regression_data.plot(kind='scatter', x='Schooling', y='Life expectancy')
plt.show()
```



GRAPH OF THE REGRESSION LINE

Figure 2

Scatterplot of life expectancy vs. years in school with trendline as shown



GRAPH OF DATA WITH CENTROID: $(\bar{x}, \bar{y}) = (12.93, 71.71)$

Figure 3

Scatterplot of life expectancy vs. years in school with centroid as shown



REFERENCES

- Jaadi, Zakaria. (2019, October 15). Everything you need to know about interpreting correlations. Medium; Towards Data Science.
<https://towardsdatascience.com/eveything-you-need-to-know-about-interpreting-correlations-2c485841c0b8>
- Rajarshi, K. (2015). Life Expectancy (WHO). Kaggle.Com.
<https://www.kaggle.com/kumarajarshi/life-expectancy-who>
-

APPENDIX A

	Country	Life expectancy	Years in School
1	Afghanistan	65	10.1
2	Albania	77.8	14.2
3	Algeria	75.6	14.4
4	Angola	52.4	11.4
5	Antigua and Barbuda	76.4	13.9
6	Argentina	76.3	17.3
7	Armenia	74.8	12.7
8	Australia	82.8	20.4
9	Austria	81.5	15.9
10	Azerbaijan	72.7	12.7
11	Bahamas	76.1	12.6
12	Bahrain	76.9	14.5
13	Bangladesh	71.8	10.2
14	Barbados	75.5	15.3
15	Belarus	72.3	15.6
16	Belgium	81.1	16.6
17	Belize	71	12.8
18	Benin	60	10.7
19	Bhutan	69.8	12.5
20	Bolivia (Plurinational State of)	77	13.8
21	Bosnia and Herzegovina	77.4	14.2
22	Botswana	65.7	12.6
23	Brazil	75	15.2
24	Brunei Darussalam	77.7	14.9

25	Bulgaria	74.5	15
26	Burkina Faso	59.9	7.7
27	Burundi	59.6	10.6
28	Cabo Verde	73.3	13.5
29	Cambodia	68.7	10.9
30	Cameroon	57.3	10.4
31	Canada	82.2	16.3
32	Central African Republic	52.5	7.1
33	Chad	53.1	7.3
34	Chile	85	16.3
35	China	76.1	13.5
36	Colombia	74.8	13.6
37	Comoros	63.5	11.1
38	Congo	64.7	11.1
39	Costa Rica	79.6	14.2
40	Croatia	78	15.3
41	Cuba	79.1	13.9
42	Cyprus	85	14.3
43	Denmark	86	19.2
44	Djibouti	63.5	6.3
45	Dominican Republic	73.9	13.2
46	Ecuador	76.2	14
47	Egypt	79	13.1
48	El Salvador	73.5	13.2
49	Equatorial Guinea	58.2	9.2
50	Eritrea	64.7	5
51	Estonia	77.6	16.5

52	Ethiopia	64.8	8.4
53	Fiji	69.9	15.3
54	Finland	81.1	17
55	France	82.4	16.3
56	Gabon	66	12.6
57	Gambia	61.1	8.9
58	Georgia	74.4	13.9
59	Germany	81	17.1
60	Ghana	62.4	11.4
61	Greece	81	17.2
62	Grenada	73.6	15.8
63	Guatemala	71.9	10.7
64	Guinea	59	8.8
65	Guinea-Bissau	58.9	9.2
66	Guyana	66.2	10.3
67	Haiti	63.5	9.1
68	Honduras	74.6	11.2
69	Hungary	75.8	15.6
70	Iceland	82.7	19
71	India	68.3	11.6
72	Indonesia	69.1	12.9
73	Iran (Islamic Republic of)	75.5	14.8
74	Iraq	68.9	10.1
75	Ireland	81.4	18.6
76	Israel	82.5	16
77	Italy	82.7	16.3
78	Jamaica	76.2	12.8

79	Japan	83.7	15.3
80	Jordan	74.1	13.1
81	Kazakhstan	72	15
82	Kenya	63.4	11.1
83	Kiribati	66.3	11.9
84	Kuwait	74.7	13.3
85	Kyrgyzstan	71.1	13
86	Lao People's Democratic Republic	65.7	10.8
87	Latvia	74.6	16
88	Lebanon	74.9	13.3
89	Lesotho	53.7	10.7
90	Liberia	61.4	9.9
91	Libya	72.7	13.4
92	Lithuania	73.6	16.5
93	Luxembourg	82	13.9
94	Madagascar	65.5	10.3
95	Malawi	58.3	10.8
96	Malaysia	75	13.1
97	Maldives	78.5	12.7
98	Mali	58.2	8.4
99	Malta	81.7	14.6
100	Mauritania	63.1	8.5
101	Mauritius	74.6	15.2
102	Mexico	76.7	13.3
103	Micronesia (Federated States of)	69.4	11.7
104	Mongolia	68.8	14.8
105	Montenegro	76.1	15.1

106	Morocco	74.3	12.1
107	Mozambique	57.6	9.1
108	Myanmar	66.6	9.1
109	Namibia	65.8	11.7
110	Nepal	69.2	12.2
111	Netherlands	81.9	18.1
112	New Zealand	81.6	19.2
113	Nicaragua	74.8	11.7
114	Niger	61.8	5.4
115	Nigeria	54.5	10
116	Norway	81.8	17.7
117	Oman	76.6	13.7
118	Pakistan	66.4	8.1
119	Panama	77.8	13
120	Papua New Guinea	62.9	10
121	Paraguay	74	12.3
122	Peru	75.5	13.4
123	Philippines	68.5	11.7
124	Poland	77.5	16.4
125	Portugal	81.1	16.6
126	Qatar	78.2	13.4
127	Romania	75	14.7
128	Russian Federation	75	15
129	Rwanda	66.1	10.8
130	Saint Lucia	75.2	13.1
131	Saint Vincent and the Grenadines	73.2	13.3
132	Samoa	74	12.9

133	Sao Tome and Principe	67.5	11.2
134	Saudi Arabia	74.5	16.1
135	Senegal	66.7	9.5
136	Serbia	75.6	14.4
137	Seychelles	73.2	14.1
138	Sierra Leone	51	9.5
139	Singapore	83.1	15.4
140	Slovakia	76.7	15
141	Slovenia	88	17.3
142	Solomon Islands	69.2	9.6
143	South Africa	62.9	13
144	South Sudan	57.3	4.9
145	Spain	82.8	17.7
146	Sri Lanka	74.9	14
147	Sudan	64.1	7.2
148	Suriname	71.6	12.7
149	Swaziland	58.9	11.4
150	Sweden	82.4	15.9
151	Switzerland	83.4	16
152	Syrian Arab Republic	64.5	9
153	Tajikistan	69.7	11.3
154	Thailand	74.9	13.6
155	The former Yugoslav republic of Macedonia	75.7	12.9
156	Timor-Leste	68.3	12.5
157	Togo	59.9	12
158	Tonga	73.5	14.3
159	Trinidad and Tobago	71.2	12.7

160	Tunisia	75.3	14.6
161	Turkey	75.8	14.5
162	Turkmenistan	66.3	10.8
163	Uganda	62.3	10
164	Ukraine	71.3	15.3
165	United Arab Emirates	77.1	13.3
166	Uruguay	77	15.5
167	Uzbekistan	69.4	12.1
168	Vanuatu	72	10.8
169	Venezuela (Bolivarian Republic of)	74.1	14.3
170	Viet Nam	76	12.6
171	Yemen	65.7	9
172	Zambia	61.8	12.5
173	Zimbabwe	67	10.3