

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

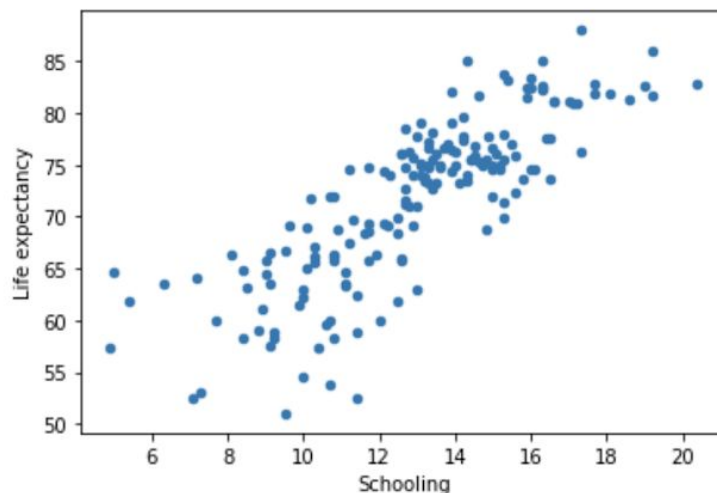
The regression line/line of best fit for life expectancy (y) vs. years of schooling (x) based from the data shown in Appendix A is  $y = 2.2287x + 42.9016$ . The slope of the line ( $b_1 = 2.2287$ ) tells us that for every increase of one year in years of schooling, we can expect an increase of 2.2287 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*

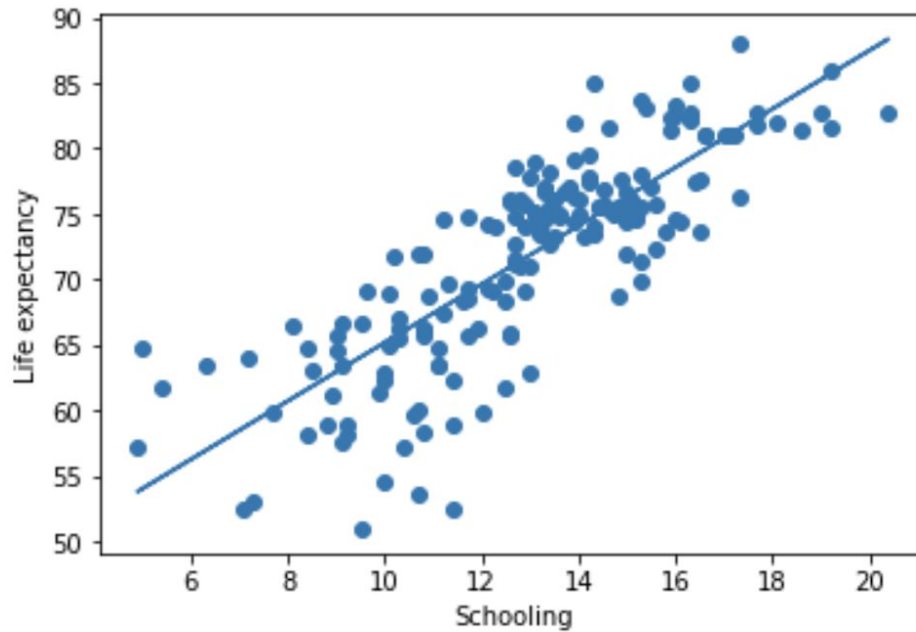
```
In [6]: regression_data.plot(kind='scatter', x='Schooling', y='Life expectancy')
plt.show()
```



## 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

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

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



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

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

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

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