SECI 2143

PROBABILITY & STATISTICAL DATA ANALYSIS

PROJECT 2

The Study of Life Expectancy, Growth Rate and Mortality Rate among Countries

LECTURER:

DR SHARIN HAZLIN BINTI HUSPI

SUBMITTED BY: GROUP 03

Felicia Chin Hui Fen (A20EC0037)

Mek Zhi Qing (A20EC0077)

Rohaizaazira Binti Mohd Zawawi (A20EC0138)

Zereen Teo Huey Huey (A20EC0173)

Semester 2 2020/2021

TABLE OF CONTENT

INTRODUCTION

The world's population continues to grow, and changes in growth rates and death rates indicate that there is a great uncertainty in its long-term development. The world population is 7.6 billion now, and it is still growing. The examples of countries with large populations are China, India, and the US while the countries with small populations are Vatican City, The Republic of Nauru, and Tuvalu (Vincej, 2020). Life expectancy is an estimate of the average number of additional years a person can expect at a given age. (Bezy, 2021) Compared with more developed countries, life expectancy in less developed countries is relatively low. In some less developed countries, due to high infant mortality, life expectancy at birth may be lower than life expectancy at 1 year. In this report, we want to study the relationship of life expectancy, mortality rate and growth rate among countries.

DATASET

The dataset that we studied is a secondary data that is presented in an excel file (CSV). The dataset is provided by our lecturer from an online resource. The sample size of the dataset is 198 countries. This dataset contains information about countries' which is country name, population, growth rate, mortality rate, life expectancy, and the number of individuals under the age of 15. We will analyze the data through a few tests which are 1-sample hypothesis test, correlation, regression, and Goodness of Fit test.

Variable	Level of Measurement
Country	Nominal
Population	Ratio
Growth Rate	Ratio
Mortality Rate	Ratio
Individual Under 15	Ratio
Life expectancy	Ratio

Table 1: Data type for variable in the dataset

DATA ANALYSIS

Hypothesis 1 Sample Test

According to the World Health Organization (WHO), the world mortality rate is 8.9 million (World Health Organization, n.d). We wish to determine whether the sample mean mortality rate is smaller than 8.90 or not at 0.05 significance level. We assume the population mean mortality rate is 8.90.

Variables used in the test is:

n = sample size
s = standard deviatio
μ is the population mean of the countries mortality rate.
$\bar{x}$ =the sample mean of the countries mortality rate.

H_0: $\mu = 8.90$ (The mean mortality rate is equal to 8.90)
H_1: $\mu < 8.90$ (The mean mortality rate is smaller than 8.90)

Significance level, $\alpha = 0.05$

The population variance is unknown. However, since the sample size is 198 which is sufficiently large, we can assume that the sample is normally distributed. Z test is applied in this test and is calculated using the R program. This is a left tailed test, thus, will reject the null hypothesis if the Z value calculated is smaller than the - Z-statistic. The Z-value calculated is -20.83545 while the - Z-statistic is -1.644854. Z-value is smaller than -Z-statistic. Thus, we will reject the null hypothesis since the Z-value calculated is fall within the critical region. As a result, at 5% significance level, we have sufficient evidence to conclude that the sample mean mortality rate is smaller than 8.90.

Correlation

The correlation test is carried out to determine whether there is linear correlation between growth rate and mortality rate. The method Pearson's product-moment correlation is used to calculate the correlation between mortality rate and growth rate among 198 countries on 5% significance level. The null hypothesis and alternative hypothesis are:

H_0: p=0 (There is no linear correlation between growth rate and mortality rate)
H_1: $p \neq 0$ (There is linear correlation between growth rate and mortality rate)

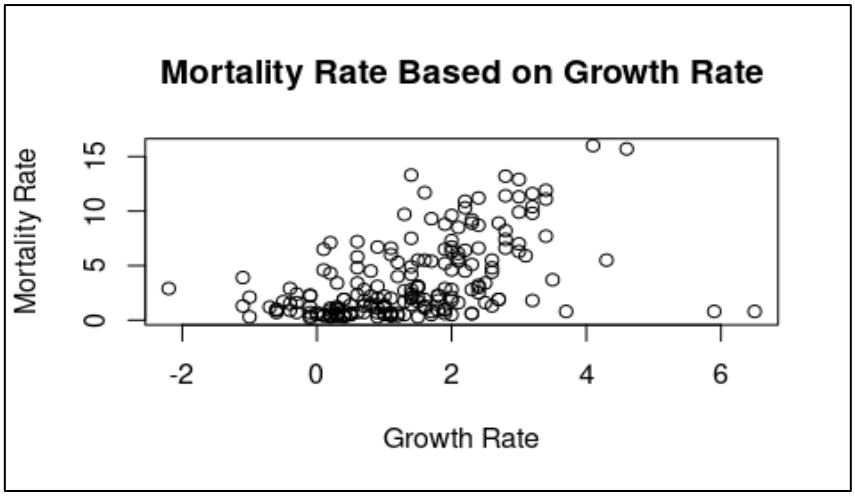

Figure 1: Scatter plot of Mortality Rate Based on Growth Rate

Based on the result calculated, t=8.2271> $t_{0.975,196}$=1.972141, hence the null hypothesis is rejected. There is sufficient evidence to conclude that there is linear correlation between growth rate and mortality rate on 5 % significance level. The correlation coefficient is +0.5066449 which indicates that there is a moderate positive linear relationship between mortality rate and growth rate. The higher the growth rate, the higher the mortality rate. The correlation can also be predicted based on the scatter plot.

Regression

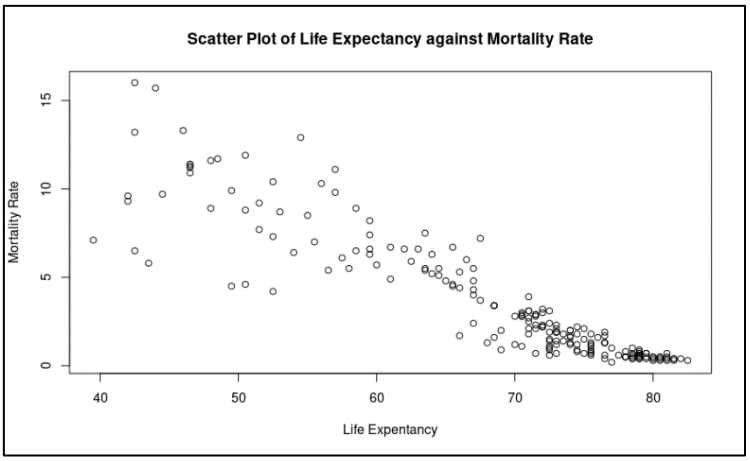

Figure 2: The scatter plot of life expectancy against mortality rate

Based on Figure 2, the regression test is carried out with the objective of determining the relationship between the life expectancy and the mortality rate of the countries given at the significance level of 0.05. The dependent variable (y) used is the mortality rate while the independent variable (x) used is the life expectancy.

H_0: $ß_1 = 0$ (There is no linear relationship between mortality rate and life expectancy)
H_1: $ß_1 \neq 0$ (There is linear relationship exist between mortality rate and life expectancy)

Through regression analysis, there exist a negative linear relationship between the life expectancy and mortality rate with the equation of:

$$\hat{y} = b_0 + b_1 x$$

where,

$\hat{y}$ = Estimated y value
x = Independent variable
b_0 = Estimate of the regression intercept = 24.1564
b_1 = Estimate of the regression slope = -0.3006

Therefore, the equation for this estimated regression model is:

$$\hat{y} = 24.1564 - 0.3006x$$

From the value of b_0 = 24.1564, it can be determined that the rate of mortality is 24.1564 when the life expectancy is 0 since it is the estimated average value of y when the value of x is zero (y-intercept). While for the value of b_1 = -0.3006, it can be indicated that the mortality rate will decrease 0.3006 on average when the life expectancy increases. Therefore, we have sufficient evidence to conclude that there exists a negative linear relationship between the life expectancy and the mortality rate of the countries given.

For the coefficient of determination, the equation shown below is required to calculate it as:

$$R^2 = SSR / SST$$

where,

R^2 = Coefficient of determination
SSR = regression sum of squares = $\sum(\hat{y} - \bar{y})^2$
SST = total sum of squares = $\sum(y - \bar{y})^2$

therefore,

$$R^2 = \sum(\hat{y} - \bar{y})^2 / \sum(y - \bar{y})^2$$
$$= 0.8299$$

Percentage of R^2 = 0.8299 x 100% = 82.99%

Therefore, it can be determined that 82.99% of the variation in mortality rate is explained by variation in life expectancy and the linear relationship between the life expectancy and the mortality rate of the countries given is quite strong since the R^2 = 0.8299 is near to 1.

Goodness of Fit Test

	Datanames	Life Expectancy
1	Afghanistan	44.0
2	Albania	76.5
3	Algeria	72.5
4	American Samoa	72.5
5	Andorra	42.5
6	Angola	75.5
7	Anguilla	71.5
8	Antarctica	74.0
9	Antigua and Barbuda	81.5
10	Arctic Ocean	80.0

Table 2: Observed Life Expectancy for the 10 Countries Data

Goodness of fit test is carried out to investigate is there any differences of the proportion of life expectancy between 10 countries data at 5 % significance level.

Variable used in the test:

n: number of tested data

observed: observed value of the life expectancy of 10 countries data

expprob: expected proportion of the life expectancy for the 10 countries data

expected: expected value of the life expectancy of 10 countries data

exp: result of square difference of the observed value minus expected value and divide by expected value,

$$\text{exp: } \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

x^2 : summation of all value of exp

Let P1 = Proportion of life expectancy in Afghanistan

P2 = Proportion of life expectancy in Albania

P3 = Proportion of life expectancy in Algeria

P4 = Proportion of life expectancy in American Samoa

P5 = Proportion of life expectancy in Andorra

P6 = Proportion of life expectancy in Angola

P7 = Proportion of life expectancy in Anguilla

P8 = Proportion of life expectancy in Antarctica

P9 = Proportion of life expectancy in Antigua and Barbuda

P10 = Proportion of life expectancy in Arctic Ocean

H_0: P1 = P2 = P3 = P4 =P5 = P6 = P7= P8 = P9 = P10 = 69.05 (The proportion of the life expectancy of the 10 countries data are the same, which is 69.05.)
H_1: At least one of the proportions of life expectancy is different from others.

Significance level, $\alpha = 0.05$
Degree of freedom = 10-1 = 9
Reject the null hypothesis if the $X^2_{stat} > X^2_{(0.05, 9)} = 16.91898$

Countries Data	Life Expectancy	Expected Life Expectancy	exp
Afghanistan	44.00	69.05	9.08765387
Albania	76.50	69.05	0.80380159
Algeria	72.50	69.05	0.17237509
American Samoa	72.50	69.05	0.17237509
Andorra	42.50	69.05	10.20858074
Angola	75.50	69.05	0.60249819
Anguilla	71.50	69.05	0.08692976
Antarctica	74.00	69.05	0.35485156
Antigua and Barbuda	81.50	69.05	2.24478639
Arctic Ocean	80.00	69.05	1.73645909

Table 3: Comparison of observed and expected life expectancy of 10 Countries Data

By using R program, we calculate the chi-square value,

$$X^2_{stat} = 25.47031$$

Since the chi-square test statistics value is 25.47031 which is greater than the chi-square statistical value, 16.91898, we reject the null hypothesis that the distribution of life expectancy for the 10 countries data are different. Thus, at 5% significance level, we have sufficient evidence to conclude that the proportion of life expectancy of these 10 countries is different from each other.

CONCLUSION

Based on all the findings through this report, a few conclusions can be made. First, the sample mean mortality rate of our dataset is lower than the world mortality rate. Next, there is a moderate relationship between growth rate and mortality rate among countries. The higher the growth rate, the higher the mortality rate. Thirdly, there is a negative linear relationship between the life expectancy and the mortality of the countries given. Lastly, for the Goodness of fit test, we can conclude that the proportion of life expectancy in every country is different. It is because of a few factors such as life standard and health standard. Each country has its own life standard and health standard. So, there are reasons why life expectancy in each country is different.

Through this study, we have learned how to use R which is a very useful tool in helping us to summarize the data. We also understand more about the topics we learnt before by applying them to solve the problems. We hope we can have the chance to learn more and apply it in our future studies.

REFERENCE

Bezy, J. (2021). life expectancy | Definition & Facts. Retrieved 15 June 2021, from https://www.britannica.com/science/life-expectancy

Vincej, V. (2020). 10 Smallest Countries in the World by Population - [2021]. Retrieved 15 June 2021, from https://www.travelinglifestyle.net/smallest-countries-world-by-population/

World Health Organization. (n.d.). *Global Health Estimates: Life expectancy and leading causes of death and disability*. Retrieved June 14, 2021, from https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates

APPENDIX

Processed Dataset

Datanames	Population	Growth	under 15	Life Expectancy	Mortality
Afghanistan	27145300	4.6	47	44	15.7
Albania	3190000	0.4	25	76.5	1.9
Algeria	33857900	1.5	28	72.5	3.1
American Samoa	57300	2.3	39	72.5	0.6
Andorra	17024100	2.8	46	42.5	13.2
Angola	39531100	1	26	75.5	1.3
Anguilla	3002300	-0.4	19	71.5	2.9
Antarctica	103900	1.5	22	74	1.7
Antigua and Barbuda	20743200	1.1	19	81.5	0.4
Arctic Ocean	8360700	0.2	15	80	0.4
Argentina	8467200	0.6	23	67.5	7.2
Armenia	331300	1.4	27	73.5	1.4
Aruba	752600	1.6	25	75.5	1.1
Ashmore and Cartier Islands	1.59E+08	1.9	34	64	5.2
Atlantic Ocean	293900	0.3	18	77	1
Australia	9688800	-0.6	15	69	0.9
Austria	10457300	0.2	17	79	0.4
Azerbaijan	287700	2.1	37	76	1.6
Bahamas	9032800	3.2	44	57	9.8
Bahrain	658500	2.2	31	65.5	4.5
Baltic Sea	9524600	2	37	65.5	4.6
Baker Island	3934800	0.3	17	74.5	1.2
Bangladesh	1881500	0.1	35	50.5	4.6
Barbados	1.92E+08	1.4	27	72.5	2.4
Bassas da India	390100	2.3	29	77.5	0.6
Belarus	7638800	-0.7	13	73	1.2
Belgium	14784300	3.2	46	52.5	10.4
Belize	8508200	3	44	49.5	9.9
Benin	14443700	2	36	59.5	6.3
Bermuda	18549200	1.9	41	50.5	8.8
Bhutan	32876000	1	17	80.5	0.5
Bolivia	530400	2.4	38	71	2.5
Borneo	4342700	1.3	42	44.5	9.7
Bosnia and Herzegovina	10780600	3.4	46	50.5	11.9
Botswana	149300	0.4	16	79	0.5
Bouvet Island	16634800	1.1	24	78.5	0.7
China	1.33E+09	0.6	21	73	2.3
British Virgin Islands	7206100	1.2	14	82	0.4
Brunei	481100	0.7	14	81	0.7
Bulgaria	46156000	1.6	29	73	1.9
Burkina Faso	839200	2.6	42	65	4.8
Burundi	3768100	3	42	55.5	7

Cambodia	18000	-1	30	71	2.1
Cameroon	4467600	1.9	27	78.5	1
Canada	19261800	1.6	41	48.5	11.7
Cape Verde	4555400	0.2	15	75.5	0.6
Cayman Islands	11267900	0.3	18	78	0.5
Central African Republic	854700	1.2	19	79	0.6
Chad	10186300	-0.1	14	76.5	0.4
Chile	23790200	0.6	23	67	4.8
China	62635700	2.8	47	46.5	11.4
Christmas Island	5442100	0.3	19	78.5	0.4
Clipperton Island	833000	2.1	37	55	8.5
Cocos Islands	9759700	1.5	33	72	3
Colombia	13341200	1.4	32	75	2.1
Comoros	75497900	1.9	33	71.5	2.9
Cook Islands	6857300	1.8	33	72	2.2
Coral Sea Islands	507500	2.3	42	51.5	9.2
Costa Rica	4850800	4.3	43	58	5.5
Cote dIvoire	1335300	-0.6	15	71.5	0.7
Croatia	83099200	2.4	44	53	8.7
Cuba	838700	0.9	32	69	2
Curacao	5276900	0.3	17	79	0.4
Cyprus	61647400	0.4	18	80.5	0.4
Czech Republic	202100	2.6	34	76.5	1.3
Democratic Republic of the Congo	262800	1.7	27	74.5	0.8
Denmark	1330600	1.7	35	56.5	5.4
Djibouti	1708700	2.8	41	59.5	7.4
Dominica	4395400	-1.1	18	71	3.9
Dominican Republic	82599500	0.1	14	79.5	0.4
East Timor	23478400	2.1	38	60	5.7
Ecuador	11146900	0.3	14	79.5	0.7
Egypt	105700	0.3	33	68.5	3.4
El Salvador	444900	0.9	24	79	0.7
Equatorial Guinea	173300	1.8	29	75.5	0.9
Eritrea	13353900	2.4	43	70.5	3
Estonia	9370100	2.2	43	56	10.3
Ethiopia	1695000	3	48	46.5	11.3
Europa Island	737900	0.2	31	67	4.3
Falkland Islands Islas Malvinas	9597900	1.4	37	61	4.9
Faroe Islands	7106000	2.3	39	70.5	2.8
Fiji	10029600	-0.3	15	73	0.7
Finland	301000	0.9	22	81.5	0.3
India	1.17E+09	1.6	32	64.5	5.5
French Guiana	2.32E+08	1.3	28	71	2.7
French Polynesia	71208400	0.9	27	71	3.1
French Southern and Antarctic Lands	28993400	2.8	41	59.5	8.2

Gabon	4300900	1.7	21	78.5	0.5
Gambia	77600	-0.1	18	77	0.2
Gaza Strip	6927700	2	28	81	0.5
Georgia	58876800	0.1	14	80.5	0.5
Germany	2713800	0.5	31	72.5	1.4
Ghana	1.28E+08	0.2	14	82.5	0.3
Gibraltar	5924200	2.7	36	72.5	1.9
Glorioso Islands	15421900	-0.3	24	67	2.4
Greece	37537700	2.2	43	54	6.4
Greenland	2851100	3.7	23	78	0.8
Grenada	5316500	1.2	30	66	5.3
Guadeloupe	5859400	2.3	38	64.5	5.1
Guam	2277000	-0.6	14	72.5	1
Guatemala	4099100	1	28	72	2.2
Guernsey	2007800	0.1	40	42.5	6.5
Guinea	3750300	1.4	47	46	13.3
Guinea-Bissau	6160500	2	30	74.5	1.8
Guyana	3389900	-0.4	16	72.5	0.9
Haiti	466600	1.3	18	79	0.5
Heard Island and McDonald Islands	19683400	2.8	43	59.5	6.6
Honduras	13925100	2.3	47	48	8.9
Hong Kong	26571900	1.9	30	74.5	0.9
Howland Island	305600	2.5	32	68.5	3.4
Hungary	12336800	3	48	54.5	12.9
Iceland	406600	0.5	17	79	0.6
India	54600	3.5	42	67.5	3.7
Indian Ocean	398700	0.5	21	79.5	0.7
Indonesia	3123800	3	40	64	6.3
Iran	1261600	1	24	73	1.4
Iraq	1.07E+08	1.3	30	76.5	1.7
Ireland	111100	0.6	38	68.5	3.4
Isle of Man	2678800	1.2	27	67	4
Israel	598000	-0.1	19	74.5	2.2
Italy	31224100	1.5	29	71	3.1
Jamaica	21396900	2	44	42	9.6
Jan Mayen	48798200	1.1	26	62	6.6
Japan	2074100	1.4	37	52.5	4.2
Jarvis Island	28196000	2.1	38	63.5	5.4
Jersey	16418800	0.5	18	80	0.5
Johnston Atoll	191600	0.8	21	75	1.5
Jordan	241700	1.9	26	76.5	0.6
Juan de Nova Island	4178500	1.1	21	80	0.5
Kazakhstan	5603200	2	37	73	2.1
Kenya	14225500	3.4	48	57	11.1
Kerguelen Archipelago	1.48E+08	2.2	44	46.5	10.9

Kingman Reef	1800	-2.2	33	70.5	2.9
Kiribati	4698100	-1	19	80.5	0.3
Kosovo	4017500	3.2	45	73.5	1.8
Kuwait	2595100	1	32	75.5	1.2
Kyrgyzstan	1.64E+08	2	36	65.5	6.7
Laos	19100	0.7	24	71	1.8
Latvia	3343400	1.8	30	75.5	1.8
Lebanon	6331000	2.1	40	57.5	6.1
Lesotho	6127100	2.4	35	72	3.2
Liberia	27902800	1.5	31	71.5	2.1
Libya	87960100	1.8	35	72	2.3
Liechtenstein	38082000	-0.1	15	75.5	0.7
Line Islands	10623000	0.5	16	78	0.5
Lithuania	3990500	0.6	21	79	0.7
Luxembourg	840600	5.9	21	75.5	0.8
Macau	48223900	0.4	18	78.5	0.4
Macedonia	3793600	-0.3	19	68.5	1.6
Madagascar	806700	1.6	26	76.5	1.3
Malawi	21437900	-0.4	15	72.5	1.5
Malaysia	1.42E+08	-0.5	15	66	1.7
Maldives	9724600	2.4	43	46.5	11.2
Mali	164900	0.8	27	74	1.3
Malta	120400	-0.1	28	71.5	2.3
Marshall Islands	187000	0.8	40	72	2.2
Martinique	24734500	2.7	34	73	1.9
Mauritania	12378500	2.4	42	63	6.6
Mauritius	9858400	-0.1	18	74	1.2
Mayotte	5865900	4.1	43	42.5	16
Mediterranean Sea	4436300	1.5	18	80	0.3
Mexico	5390000	0	16	75	0.7
Micronesia	2001500	0	14	78	0.5
Midway Islands	495700	2.6	40	63.5	5.5
Moldova	8698500	3.2	44	48	11.6
Monaco	48576800	0.8	32	49.5	4.5
Mongolia	44279200	1.1	15	81	0.4
Montenegro	19299200	0.9	23	72.5	1.1
Montserrat	38560500	1.9	40	58.5	6.5
Morocco	458000	0.7	29	70.5	2.8
Mozambique	1141400	0.2	39	39.5	7.1
Myanmar	9119000	0.4	17	81	0.3
Namibia	7484000	0.2	16	81.5	0.4
Nauru	19928500	2.5	36	74	1.6
Navassa Island	6736000	1.1	38	66.5	6
Nepal	63883700	0.2	21	70.5	1.1
Netherlands	1154800	0.9	45	61	6.7
New Caledonia	6585100	2.7	43	58.5	8.9

New Zealand	100300	0.4	37	73	1.9
Nicaragua	1333300	0.3	21	70	1.2
Niger	10327300	1.1	25	74	2
Nigeria	74876700	1.4	27	71.5	2.8
Niue	4965300	1.4	30	63.5	7.5
Norfolk Island	30883800	3.4	49	51.5	7.7
North Korea	46205400	-1.1	14	68	1.3
North Sea	4380400	6.5	20	79	0.8
Northern Mariana Islands	60768900	0.3	18	79.5	0.5
Norway	40453500	2	44	52.5	7.3
USA	3.06E+08	1	20	78.5	0.6
Pacific Ocean	111400	0.2	23	79	0.9
Pakistan	3339700	0.7	23	76.5	1.3
Palau	27372300	1.5	32	67	5.5
Palmyra Atoll	226200	2	39	70	2.8
Panama	87375200	1.4	28	74	2
Papua New Guinea	480000	2.6	30	66	4.4
Paracel Islands	22389200	3.1	45	62.5	5.9
Paraguay	11922000	1.7	46	42	9.3
Peru	13349400	0.6	38	43.5	5.8