Statistics Department,

Fergusson College,

(Established 1885)

FACTORS AFFECTING LIFE EXPECTANCY OF A COUNTRY

PROJECT BY:

AJMAL ANSARI

Year 2013-14
F. Y. B. Sc. Statistics

CERTIFICATE

This is to certify that the project report entitled **"FACTORS AFFECTING LIFE EXPECTANCY OF A COUNTRY"** is an original work done by **Ajmal Ansari** under my supervision and able guidance.

Prof. Mr Prashant Kulkarni,
Project Guide,
Dept. of Statistics,
Fergusson College,
Pune.

Prof. S.S. Shende
Head of Dept.
Dept. of Statistics,
Fergusson College,
Pune.

BACKGROUND

DEFINITION OF LIFE EXPECTANCY:

Life expectancy is the expected (in the statistical sense) number of years of life remaining at a given age.[1] It is denoted by e_x,[a] which means the average number of subsequent years of life for someone now aged x, according to a particular mortality experience. Because life expectancy is an average, a particular person may well die many years before or many years after their "expected" survival. The term "maximum life span" has a quite different meaning. The "median life span" is also a different concept although fairly similar to life expectancy numerically in most developed countries.

INTERPRETATION:

It is important to note that life expectancy is an average. In many cultures, particularly before modern medicine was widely available, the combination of high infant mortality and deaths in young adulthood from accidents, epidemics, plagues, wars, and childbirth, significantly lowers the overall life expectancy. But for someone who survived past these early hazards, living into their sixties or seventies would not be uncommon. For example, a society with a life expectancy of 40 may have very few people dying at age 40: most will die before 30 years of age or after 55.

In countries with high infant mortality rates, life expectancy at birth is highly sensitive to the rate of death in the first few years of life. Because of this sensitivity to infant mortality, simple life expectancy at age zero can be subject to gross misinterpretation, leading one to believe that a population with a low overall life expectancy will necessarily have a small proportion of older people.[4] For example, in a hypothetical stationary population in which half the population dies before the age of five, but everybody else dies at exactly 70 years old, the life expectancy at age zero will be about 37 years, while about 25% of the population will be between the ages of 50 and 70. Another measure such as life expectancy at age 5 (e5) can be used to exclude the effect of infant mortality to provide a simple measure of overall mortality rates other than in early childhood—in the hypothetical population above, life expectancy at age 5 would be another 64 years. Aggregate population measures such as the proportion of the population in various age classes should also be used alongside individual-based measures like formal life expectancy when analyzing population structure and dynamics.

CALCULATING LIFE EXPECTANCIES:

The starting point for calculating life expectancies is the age-specific death rates of the population members. In the past, a very simple model of age-specific mortality used the Gompertz function, although these days more sophisticated methods are used. In cases where the amount of data is relatively small, the most common methods are to fit a mathematical formula, such as an extension of the Gompertz function, to the data, or to look at an established mortality table previously derived for a larger population and make a simple adjustment to it (e.g. multiply by a constant factor) to fit the data.

With a large amount of data, one looks at the mortality rates actually experienced at each age, and applies smoothing (e.g. by cubic splines) to iron out any apparently random statistical fluctuations from one year of age to the next.

While the data required are easily identified in the case of humans, the computation of life expectancy of industrial products and wild animals involves more indirect techniques. The life expectancy and demography of wild animals are often estimated by capturing, marking and recapturing them. The life of a product, more often termed shelf life is also computed using similar methods. In the case of long-lived components such as those used in critical applications, such as in aircraft methods such as accelerated aging are used to model the life expectancy of a component. The age-specific death rates are calculated separately for separate groups of data which are believed to have different mortality rates (e.g. males and females, and perhaps smokers and non-smokers if data is available separately for those groups) and are then used to calculate a life table, from which one can calculate the probability of surviving to each age. In actuarial notation the probability of surviving from age x to age $x+n$ is denoted ${}_np_x$ and the probability of dying during age x (i.e. between ages x and $x+1$) is denoted q_x. For example, if 10% of a group of people alive at their 90th birthday die before their 91st birthday, then the age-specific death probability at age 90 would be 10%. Note that this is a probability rather than a mortality rate.

The expected future lifetime of a life age x in whole years (the *curtate expected lifetime* of *(x)*) is denoted by the symbol e_x.[a] It is the conditional expected future lifetime (in whole years), assuming survival to age x. If $K(x)$ denotes the curtate future lifetime at x, then

$$e_x = E[K(x)] = \sum_{k=0}^{\infty} k \, Pr(K(x) = k) = \sum_{k=0}^{\infty} k \; {}_kp_x \; q_{x+k}.$$

Substituting ${}_kp_x \, q_{x+k} = {}_kp_x - {}_{k+1}p_x$ in the sum and simplifying, we get the equivalent formula

$$e_x = \sum_{k=1}^{\infty} {}_kp_x.$$

If we make the assumption that on average people live a half year in the year of death, then the complete expectation of future lifetime at age x is $e_x + 1/2$.

Life expectancy is by definition an arithmetic mean. It can also be calculated by integrating the survival curve from ages 0 to positive infinity (or equivalently to the maximum lifespan, sometimes called 'omega'). For an extinct or completed cohort (all people born in year 1850, for example), of course, it can simply be calculated by averaging the ages at death. For cohorts with some survivors, it is estimated by using mortality experience in recent years. These estimates are called period cohort life expectancies.

It is important to note that this statistic is usually based on past mortality experience, and assumes that the same age-specific mortality rates will continue into the future. Thus such life expectancy figures need to be adjusted for temporal trends before calculating how long a currently living individual of a particular age is expected to live. Period life expectancy remains a commonly used statistic to summarize the current health status of a population.

However for some purposes, such as pensions calculations, it is usual to adjust the life table used, thus assuming that age-specific death rates will continue to decrease over the years, as they have done in the past. This is often done by simply extrapolating past trends; however some models do exist to account for the evolution of mortality (e.g., the Lee–Carter model).

As discussed above, on an individual basis, there are a number of factors that have been shown to correlate with a longer life. Factors that are associated with variations in life expectancy include family history, marital status, economic status, physique, exercise, diet, drug use including smoking and alcohol consumption, disposition, education, environment, sleep, climate, and health care.

LIFE EXPECTANCY FORECASTING:

Forecasting life expectancy and mortality forms an important subdivision of demography. Future trends in life expectancy have huge implications for old-age support programs like U.S. Social Security and pension systems, because the cash flow in these systems depends on the number of recipients still living (along with the rate of return on the investments or the tax rate in PAYGO systems). With longer life expectancies, these systems see increased cash outflow; if these systems underestimate increases in life-expectancies, they won't be prepared for the large payments that will inevitably occur as humans live longer and longer.

Life expectancy forecasting usually is based on two different approaches:

- Forecasting the life expectancy directly, generally using ARIMA or other time series extrapolation procedures: This approach has the advantage of simplicity, but it cannot account for changes in mortality at specific ages, and the forecasted number cannot be used to derive other life table results. Analyses and forecasts using this approach can be done with any common statistical/ mathematical software package, like EViews, R, SAS, Stata, Matlab, or SPSS.
- Forecasting age specific death rates and computing the life expectancy from the results with life table methods: This approach is usually more complex than simply forecasting life expectancy because the analyst must deal with correlated age specific mortality rates, but it seems to be more robust than simple one dimensional time series approaches. This approach also yields a set of age specific rates that may be used to derive other measures, like survival curves or life expectancies at different ages. The most important approach within this group is the Lee-Carter model,[70] which uses the singular value decomposition on a set of transformed age-specific mortality rates to reduce their dimensionality to a single time series, forecasts that time series, and then recovers a full set of age-specific mortality rates from that forecasted value. Software for this approach include Professor Rob J. Hyndman's R package called `demography` and UC Berkeley's LCFIT system.

POLICY USES OF LIFE EXPECTANCY:

- Life expectancy is one of the factors in measuring the Human Development Index (HDI) of each nation, along with adult literacy, education, and standard of living.[71]
- Life expectancy is also used in describing the physical quality of life of an area.
- Disparities in life expectancy are often cited as demonstrating the need for better medical care or increased social support.
- Life expectancies are also used when determining the value of a life settlement, a life insurance policy sold for a cash asset.

List of Countries by Life Expectancy

SAMPLE DATA							
Sr. No.	Overall Rank	Country	Overall Life Expectancy	Male Life Expectancy	Male Rank	Female Life Expectancy	Female Rank
1	1	Japan	83	79	12	86	1
2	4	Andorra	82	79	12	85	2
3	17	Austria	81	78	24	84	9
4	27	Belgium	80	78	24	83	20
5	33	Bahrain	79	78	24	80	42
6	38	Barbados	78	75	45	80	42
7	42	Brunei Darussalam	77	76	36	78	57
8	49	Argentina	76	72	67	79	51
9	61	Antigua and Barbuda	75	73	56	76	87
10	73	Albania	74	73	56	75	102
11	92	Dominican Republic	73	72	67	73	115
12	98	Cape Verde	72	68	108	76	87
13	108	Armenia	71	67	116	75	102
14	115	Bangladesh	70	69	103	70	128
15	118	Guatemala	69	66	122	73	115
16	126	Laos	68	66	122	69	131
17	131	Bhutan	67	66	122	69	131
18	136	Botswana	66	64	139	67	138
19	138	Cambodia	65	64	139	66	143
20	143	Ghana	64	62	149	65	146
21	147	Guyana	63	60	151	67	138
22	152	Comoros	62	60	151	63	155
23	155	Eritrea	61	59	156	64	150
24	157	Afghanistan	60	59	156	61	158
25	162	Liberia	59	58	160	60	162
26	165	Gambia	58	57	165	60	162
27	170	Benin	57	56	171	59	168
28	171	Burkina Faso	56	54	175	57	173
29	176	Guinea	55	54	175	56	176

Continent wise Classification of Countries

SAMPLE DATA				
Sr. No.	Name of Continent	Countries	Life Expectancy	Mean
1	Asia	Japan	83	71.66666667
2		Bahrain	79	
3		Sri Lanka	75	
4		Thailand	74	
5		Tanzania	59	
1	Africa	Mauritius	74	58.38
2		Madagascar	66	
3		Sudan	62	
4		Mauritania	59	
5		Sierra Leone	47	
1	North America	Canada	82	74.63636364
2		Barbados	78	
3		Bahamas	75	
4		Nicaragua	73	
5		Haiti	63	
1	South America	Chile	79	74.08333333
2		Peru	77	
3		Ecuador	76	
4		Brazil	74	
5		Guyana	63	
1	Europe	San Marino	83	77.82222222
2		Luxembourg	82	
3		Greece	81	
4		Slovenia	80	
5		Russia	69	
1	Australia and Oceania	Australia	82	70.9375
2		Cook Islands	77	
3		Palau	72	
4		Nauru	71	

Relation between GDP per capita and Life Expectancy of a Country

SAMPLE DATA				
Sr. No.	Rank	Country	Intl. $	Life Expectancy
1	1	Qatar	102,943	82
2	5	Brunei	49,384	77
3	9	Netherlands	42,183	81
4	14	Australia	40,234	82
5	19	Taiwan	37,720	80
6	24	Japan	34,740	83
7	28	Spain	30,626	82
8	33	Bahrain	27,556	79
9	38	Seychelles	24,726	74
10	43	Estonia	20,380	76
11	48	Lithuania	18,856	74
12	53	Russia	16,736	69
13	58	Malaysia	15,568	74
14	63	Mexico	14,610	75
15	68	Bulgaria	13,597	74
16	73	Romania	12,476	74
17	—	World[5]	11,489	75
18	82	Azerbaijan	10,202	71
19	87	Dominican Republic	9,287	73
20	92	China	8,382	76
21	97	El Salvador	7,550	72
22	102	Ukraine	7,233	71
23	107	Jordan	5,900	74
24	112	Georgia	5,491	72
25	117	Morocco	5,052	72
26	122	Indonesia	4,666	70
27	127	Cape Verde	3,947	72
28	132	Vietnam	3,359	75
29	137	Pakistan	2,787	67

DESCRIPTIVE STATISTICS OF THE GIVEN DATA

Male Life Expectancy

SUMMARY STATISTICS		
Sr. No.	Property	Value
1	Mean	67.79274611
2	Standard Error	0.6332084
3	Median	70
4	Mode	72
5	Standard Deviation	8.796812232
6	Sample Variance	77.38390544
7	Kurtosis	-0.476513008
8	Skewness	-0.560900884
9	Range	37
10	Minimum	46
11	Maximum	83
12	Sum	13084
13	Count	193

Female Life Expectancy

SUMMARY STATISTICS		
Sr. No.	Property	Value
1	Mean	72.48704663
2	Standard Error	0.69889612
3	Median	76
4	Mode	78
5	Standard Deviation	9.709375199
6	Sample Variance	94.27196675
7	Kurtosis	-0.458667593
8	Skewness	-0.763723294
9	Range	39
10	Minimum	47
11	Maximum	86
12	Sum	13990
13	Count	193

Overall Life Expectancy

SUMMARY STATISTICS		
Sr. No.	Properties	Values
1	Mean	70.11398964
2	Standard Error	0.661672224
3	Median	73
4	Mode	74
5	Standard Deviation	9.192244305
6	Sample Variance	84.49735535
7	Kurtosis	-0.444733331
8	Skewness	-0.690031347
9	Range	36
10	Minimum	47
11	Maximum	83
12	Sum	13532
13	Count	193

SOME OTHER PROPERTIES OF THE GIVEN DATA

(FOR OVERALL LIFE EXPECTANCY)

Sr. No.	Property	Value
1	Geometric Mean	69.46612
2	Harmonic Mean	68.76792846
3	Trimmed Mean (10%)	70.54285714
4	Simple Mean	70.11398964

Interpretation:

- *The probability distribution curve of Life Expectancy is negatively skewed and platykurtic.(As* $\boldsymbol{\gamma_1}$=-0.69 & $\boldsymbol{\gamma}_2$ =-0.44)
- *The Mean of Female Life Expectancy is greater than Mean of Male Life Expectancy.*
- *The Maximum and Minimum values of Female Life Expectancy are greater than those of Male Life Expectancy.*
- *The Spread of Female Life Expectancy is more than that of Male Life Expectancy as the Range of Female Life Expectancy is greater than that of Male Life Expectancy.*

List of Continents By Life Expectancy

Rank	Continents	Mean Life Expectancy	Country Having the Highest Life Expectancy
1	Europe	77.82222222	San Marino(83)
2	North America	74.63636364	Canada(82)
3	South America	74.08333333	Chile(79)
4	Asia	71.66666667	Japan(83)
5	Australia and Oceania	70.9375	Australia(82)
6	Africa	58.38	Mauritius(74)

Calculation of Weighted Arithmetic Mean of Life Expectancies of Continents

Sr. No.	Continents	Mean Life Expectancy(x)	Normalized weight(w)	w*x
1	Asia	71.66666667	0.248704663	17.8238342
2	Africa	58.38	0.259067358	15.12435233
3	North America	74.63636364	0.113989637	8.507772021
4	South America	74.08333333	0.062176166	4.606217617
5	Europe	77.82222222	0.233160622	18.14507772
6	Australia and Oceania	70.9375	0.082901554	5.880829016
			Weighted A.M=70.088	

Interpretation:

- *The Maximum Mean Life Expectancy is found out to be of the Europe continent.*
- *Africa is the continent with Minimum Life Expectancy.*
- *The Weighted Arithmetic Mean of Life Expectancies of Different continents is found out to be* 70.088.

CONSISTENCY OF THE LIFE EXPECTANCY IN A CONTINENT

COMPARISON OF CONSISTENCIES OF LIFE EXPECTANCY OF DIFFERENT CONTINENTS				
Sr. No.	Name Of Continent	Mean of Life Expectancies	Standard Deviation of Life Expectancies	Coefficient of Variation
1	Asia	72.13157895	5.938340584	8.23265021
2	Africa	68.85	6.943515788	10.0849903
3	North America	74.63636364	3.711719211	4.97307081
4	South America	74.08333333	4.499228329	6.07319909
5	Europe	77	3.883233984	5.04316102
6	Australia and Oceania	76.1666667	5.706013823	7.49148423

Interpretation:

- *The Coefficient of Variation is the lowest for the Continent North America. Hence it is more stable than any other Continent mentioned above.*
- *The Coefficient of Variation is the highest for the Continent Africa. Hence it is less stable than any other Continent mentioned above.*
- *The Mean Life Expectancies of the Continents North America & South America are approximately the same as well as the Coefficients of Variation of both the above mentioned Continents are having the same value. This implies that the above mentioned two continents behave as the Clusters of the same population i.e. America.*
- *Note: The Mutual Variation among the Clusters chosen in a population must be as small as possible is a necessary requirement for Cluster Sampling. Hence the two Continents namely the North America and the South America can be treated as Clusters of a common population i.e. America.*

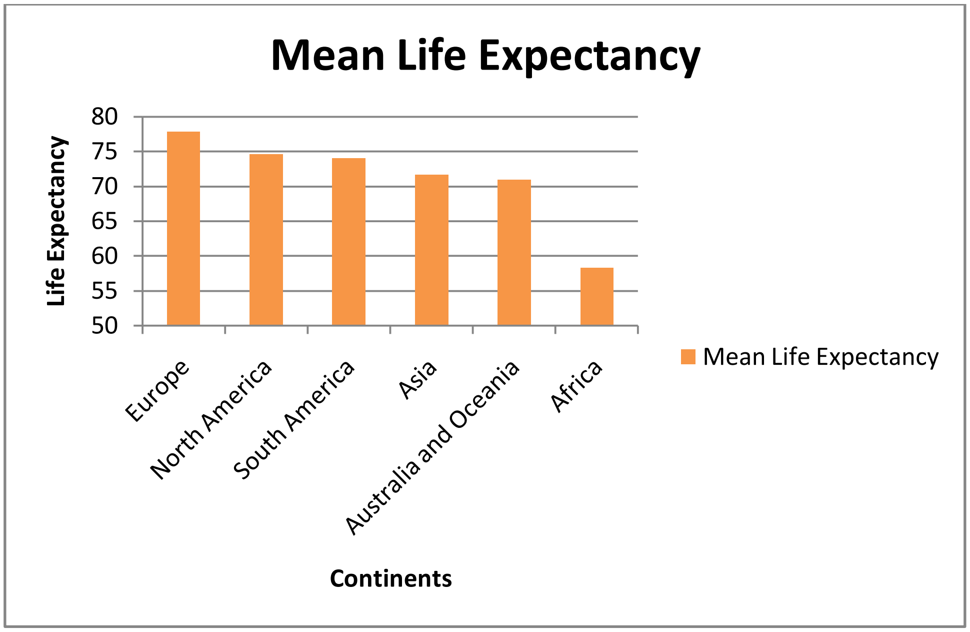
Calculation Of Weighted Arithmetic Mean Of Life Expectancies Of Countries

(SAMPLE DATA)

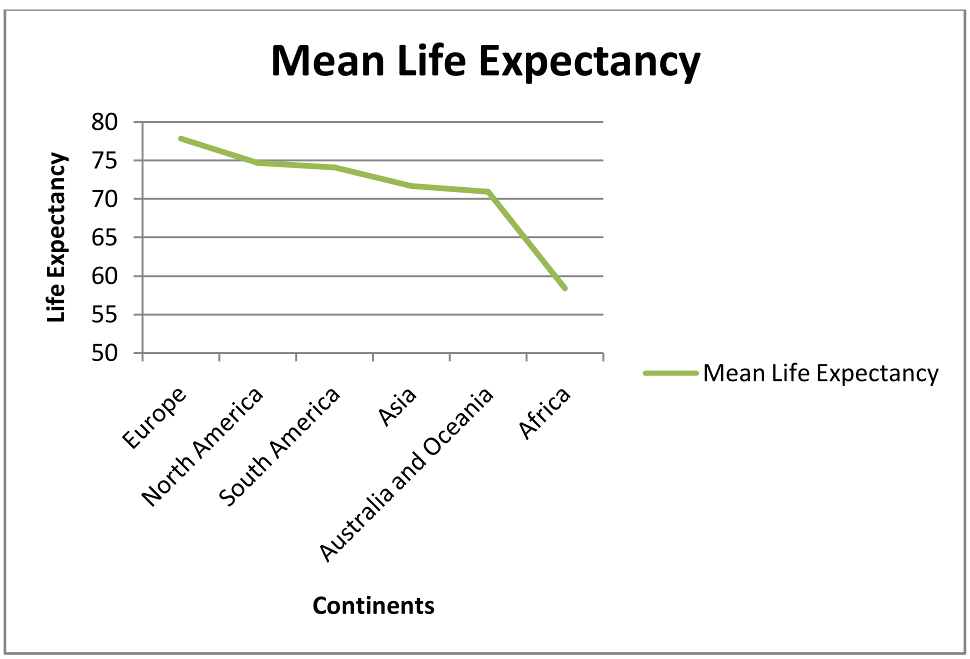
Sr. No.	Rank	Country / Territory	Population	% of World population	Normalized Weights (w)	Life Expectancy(x)	Product (w*x)
1	6	Pakistan	171,773,000	2.49%	0.0249	67	1.6683
2	12	Philippines	94,013,200	1.36%	0.0136	69	0.9384
3	18	Turkey	73,722,988	1.07%	0.0107	76	0.8132
4	24	South Africa	49,991,300	0.72%	0.0072	58	0.4176
5	30	Tanzania	43,187,823	0.63%	0.0063	59	0.3717
6	36	Morocco	32,055,000	0.46%	0.0046	72	0.3312
7	42	Nepal	28,584,975	0.41%	0.0041	68	0.2788
8	48	Australia	22,556,000	0.32%	0.0032	82	0.2624
9	54	Sri Lanka	20,410,000	0.30%	0.003	75	0.225
10	60	Kazakhstan	16,433,000	0.24%	0.0024	67	0.1608
11	66	Ecuador	14,325,000	0.21%	0.0021	76	0.1596
12	72	Greece	11,306,183	0.16%	0.0016	81	0.1296
13	78	Bolivia	10,426,154	0.15%	0.0015	67	0.1005
14	84	Belarus	9,483,100	0.14%	0.0014	71	0.0994
15	90	Austria	8,396,760	0.12%	0.0012	81	0.0972
16	96	Papua New Guinea	6,888,000	0.10%	0.001	63	0.063
17	102	El Salvador	6,194,000	0.09%	0.0009	72	0.0648
18	108	Eritrea	5,224,000	0.08%	0.00076	61	0.04636
19	114	Central African Republic	4,506,000	0.07%	0.00065	48	0.0312
20	120	Liberia	4,102,000	0.06%	0.00059	59	0.03481
21	126	Armenia	3,254,300	0.05%	0.00047	71	0.03337
22	132	Oman	2,694,094	0.04%	0.00039	72	0.02808
23	138	Botswana	1,800,098	0.03%	0.00026	66	0.01716
24	144	Trinidad and Tobago	1,317,714	0.02%	0.00019	71	0.01349
25	150	Bahrain	807,000	0.01%	0.00012	79	0.00948
					Total=0.9781		Total=69.26

Interpretation: *The Calculated value of Weighted Mean Life Expectancy is 69.2631*.

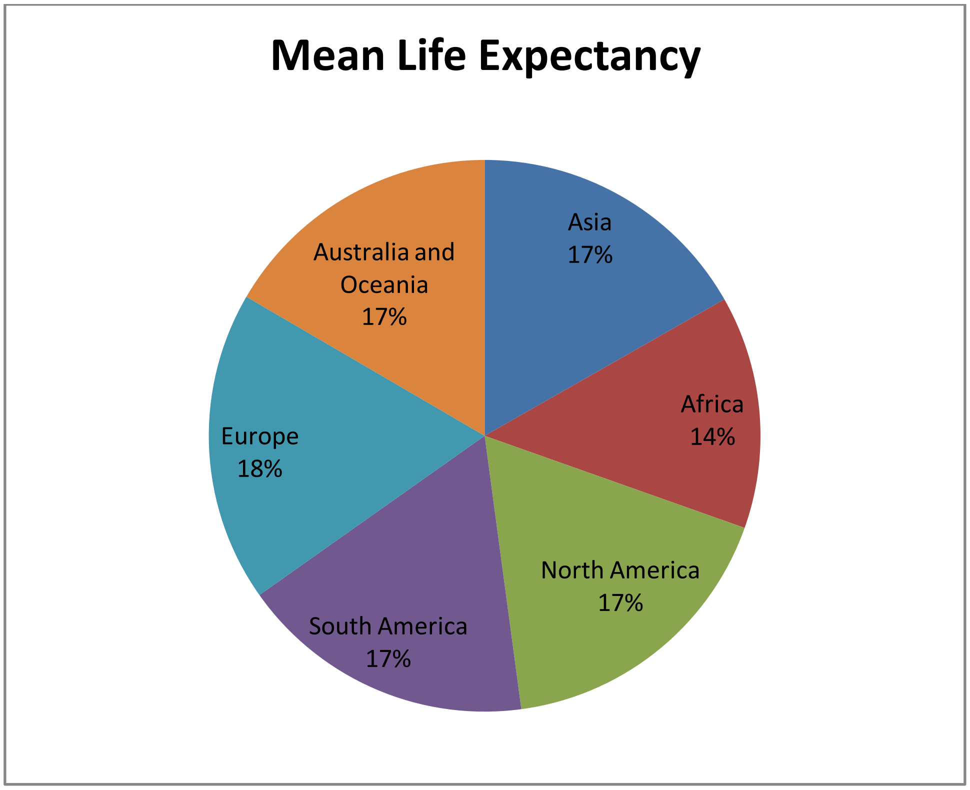
VERTICAL BAR GRAPH



LINE GRAPH



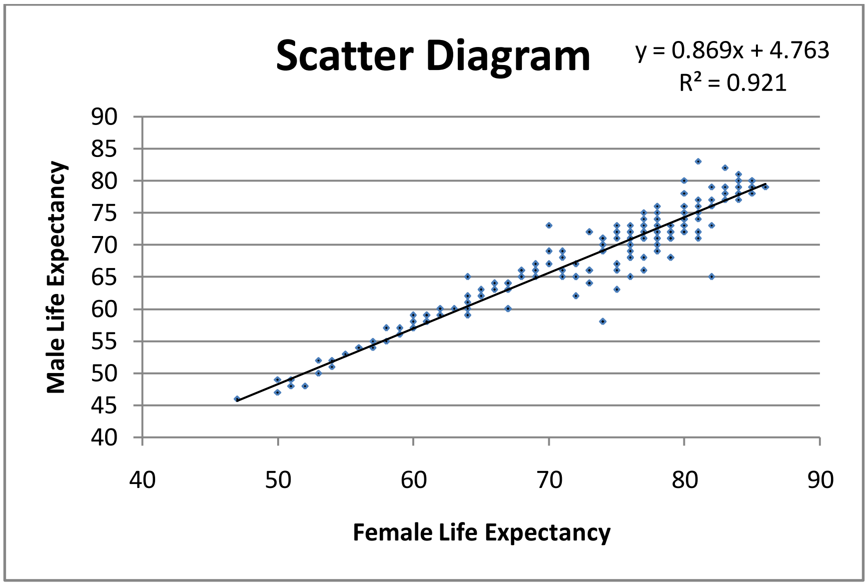
PIE DIAGRAM



Interpretation:

- *The <u>Maximum Mean Life Expectancy</u> is found out to be of the <u>Europe</u> continent.*
- *<u>Africa</u> is the continent with <u>Minimum Life Expectancy</u>.*
- *The Continents Asia, Australia & Oceania, North America and South America have the same Mean Life Expectancy approximately.*

SCATTER DIAGRAM



CORRELATION=0.958721
COVARIANCE=81.54654

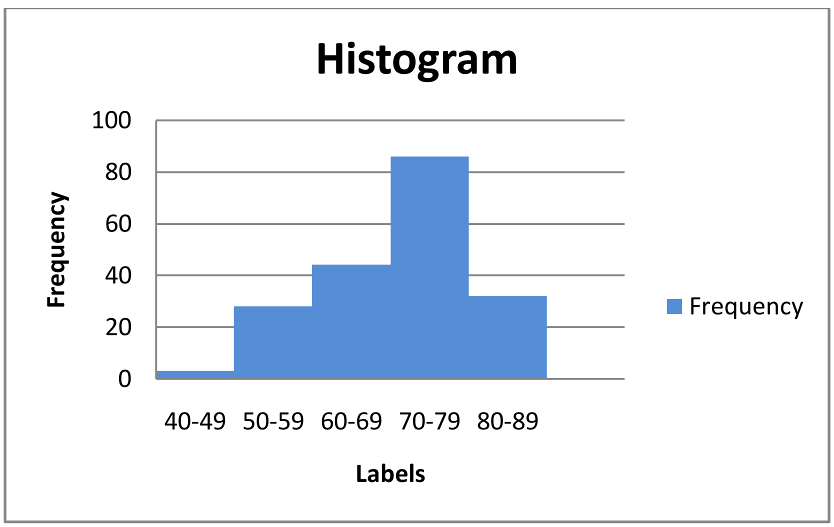
Interpretation:

- *There is a* *<u>Strong Positive type of Correlation</u>* *between the Male and female Life Expectancy*.

FEQUENCY DISTRIBUTION AND HISTOGRAM

(FOR OVERALL LIFE EXPECTANCY ONLY)

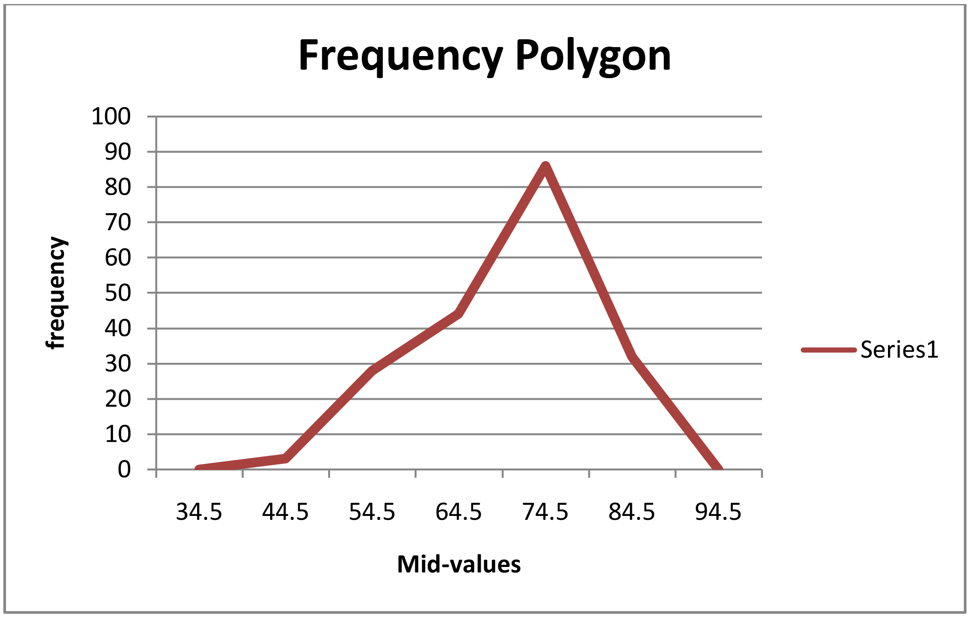
UPPER LIMITS	*Frequency*	Labels
49	3	40-49
59	28	50-59
69	44	60-69
79	86	70-79
89	32	80-89
More	0	
	SUM=193	

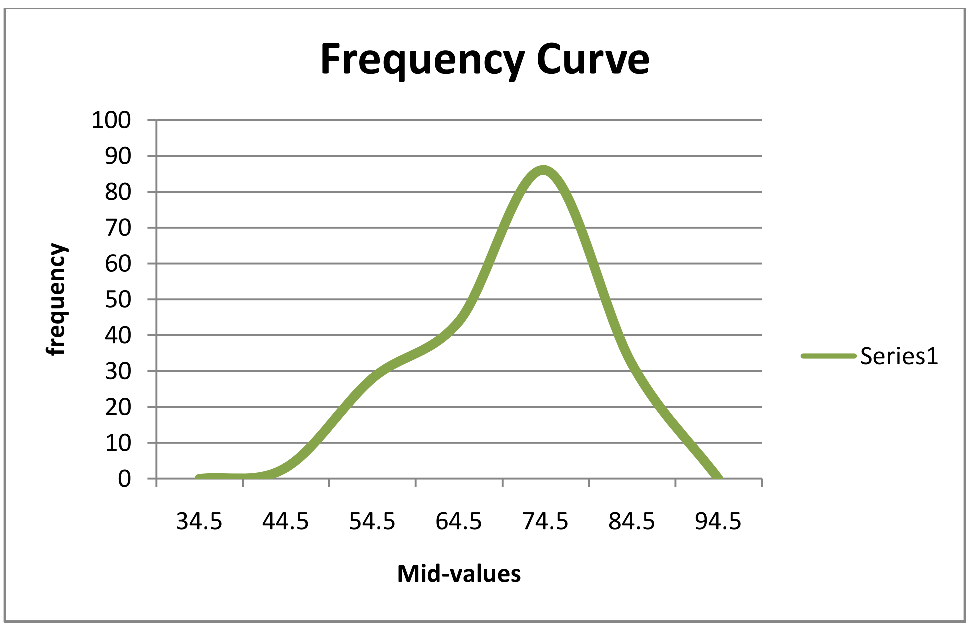


Interpretation:

- *The Modal Class observed is* 70-79.

FEQUENCY POLYGON AND FREQUENCY CURVE



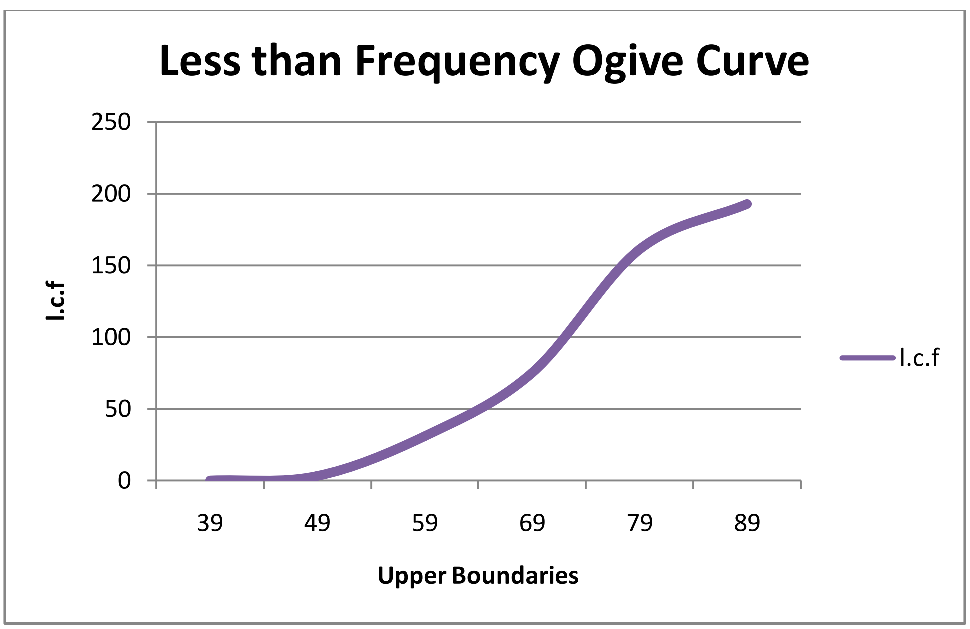


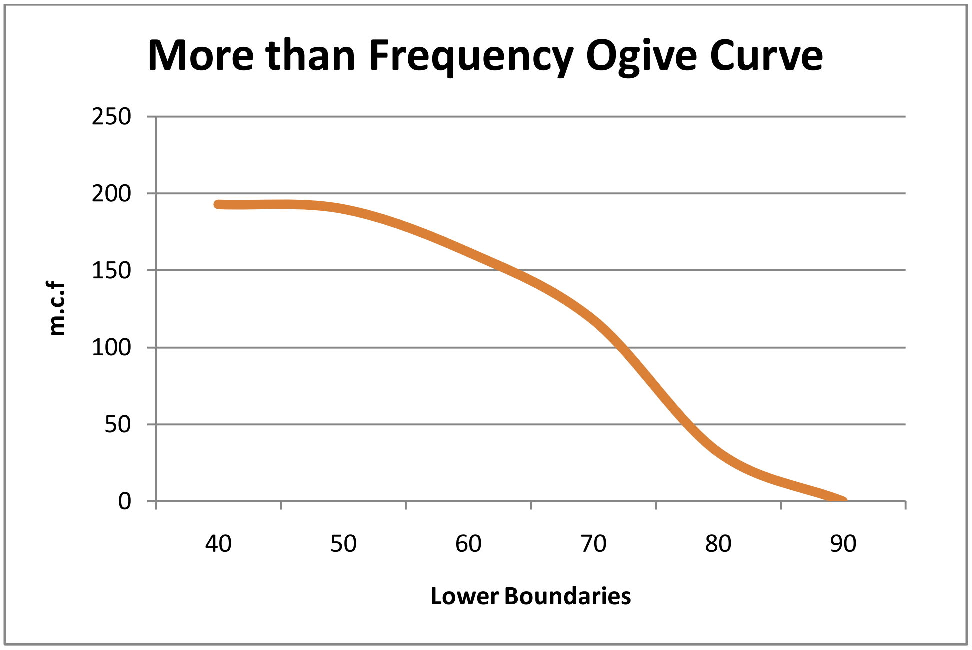
Interpretation:

- *The Highest Frequency is observed for the mid value of* 74.5~74.*Hence the* *Mode of the distribution is* 74.

LESS THAN AND MORE THAN FREQUENCY OGIVE CURVES

Classes	Frequency	l.c.f	m.c.f	Upper boundaries	Lower boundaries
40-49	3	3	193	49	40
50-59	28	31	190	59	50
60-69	44	75	162	69	60
70-79	86	161	118	79	70
80-89	32	193	32	89	80





Interpretations:

- *The less than frequency ogive curve is slightly* <u>*S-shaped*</u> *curve.*
- *The more than frequency ogive curve is* <u>*inverted S-shaped*</u>*.*

REGRESSION ANALYSIS

(FOR MALE AND FEMALE LIFE EXPECTANCY)

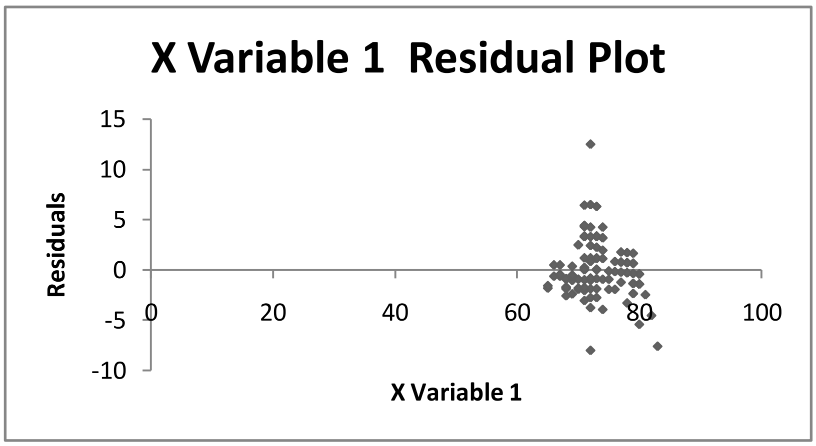
Let variable X: Female Life Expectancy (Independent variable)

& variable Y: Male Life Expectancy (Study variable)

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.95972101
R Square	0.92106442
Adjusted R Square	**0.92065115**
Standard Error	2.73502697
Observations	193

RESIDUAL PLOT



Interpretation:

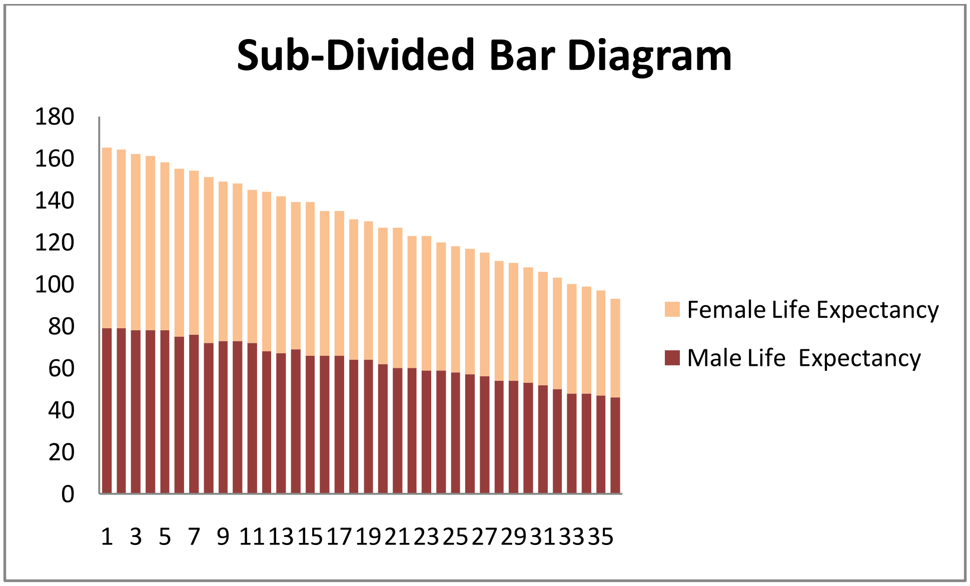
- *There is a <u>Strong Positive type of Correlation</u> between the Male and Female Life Expectancy.*
- *The <u>Coefficient of Determination</u> (R Square) obtained is <u>0.92106442</u>. Hence approximately <u>92% of variation</u> in the Study variable is explained by regression.*
- *The <u>Residual Plot</u> between the variable X and residuals <u>does not show any Specific type of Pattern</u> & hence the Regression fitted is <u>good</u>.*

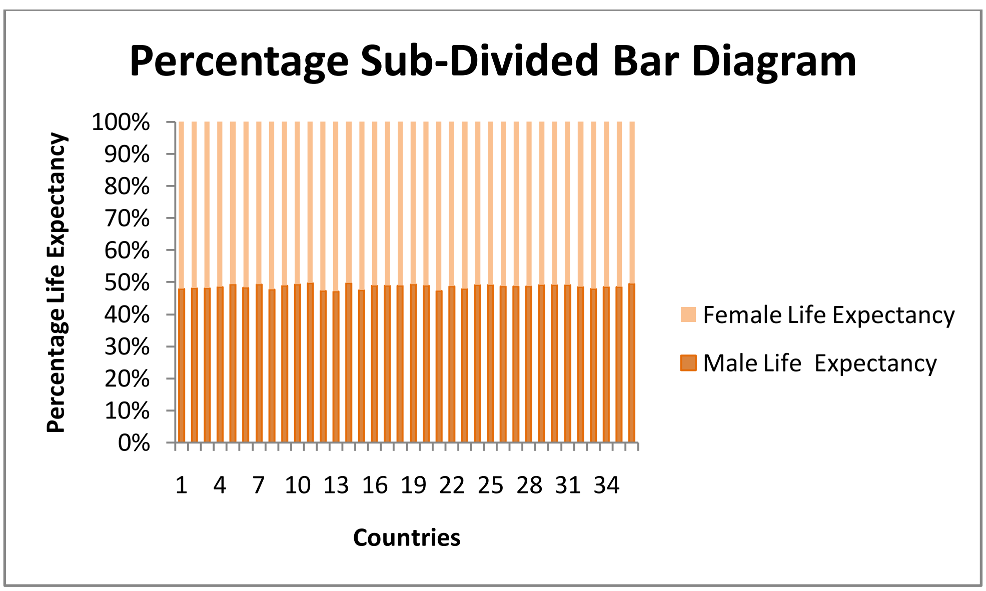
RESIDUAL OUTPUT

(SAMPLE DATA)

Observation	*Predicted Y*	*Residuals*	*Observation*	*Predicted Y*	*Residuals*
1	84.3586718	1.641328206	31	82.24011085	0.759889145
2	87.5365132	-4.536513204	32	84.35867179	-2.358671794
3	85.4179523	-0.417952264	33	83.29939132	-3.299391325
4	84.3586718	0.641328206	34	81.18083038	0.819169615
5	85.4179523	-1.417952264	35	82.24011085	-1.240110855
6	85.4179523	-1.417952264	36	82.24011085	-0.240110855
7	83.2993913	1.700608675	37	81.18083038	-0.180830385
8	86.4772327	-2.477232734	38	80.12154991	-0.121549915
9	85.4179523	-1.417952264	39	79.06226945	1.937730555
10	85.4179523	-0.417952264	40	81.18083038	-1.180830385
11	84.3586718	-0.358671794	41	80.12154991	0.878450085
12	84.3586718	0.641328206	42	81.18083038	-3.180830385
13	88.5957937	-7.595793674	43	78.00298898	3.997011025
14	85.4179523	-0.417952264	44	79.06226945	0.937730555
15	84.3586718	0.641328206	45	81.18083038	-3.180830385
16	85.4179523	-1.417952264	46	79.06226945	0.937730555
17	83.2993913	0.700608675	47	80.12154991	-2.121549915
18	84.3586718	-0.358671794	48	78.00298898	1.997011025
19	83.2993913	0.700608675	49	76.94370851	2.056291495
20	83.2993913	-0.299391325	50	79.06226945	-1.062269445
21	83.2993913	0.700608675	51	79.06226945	-2.062269445
22	84.3586718	-1.358671794	52	78.00298898	0.997011025
23	84.3586718	-1.358671794	53	75.88442804	5.115571965
24	84.3586718	-1.358671794	54	78.00298898	-0.002988975
25	84.3586718	-1.358671794	55	76.94370851	4.056291495
26	82.2401109	1.759889145	56	79.06226945	0.937730555
27	83.2993913	-0.299391325	57	76.94370851	3.056291495
28	85.4179523	-5.417952264	58	79.06226945	-1.062269445
29	84.3586718	-2.358671794	59	78.00298898	-0.002988975
30	82.2401109	0.759889145	60	80.12154991	-3.121549915

SUB-DIVIDED BAR DIAGRAM AND PERCENTAGE BAR DIAGRAM

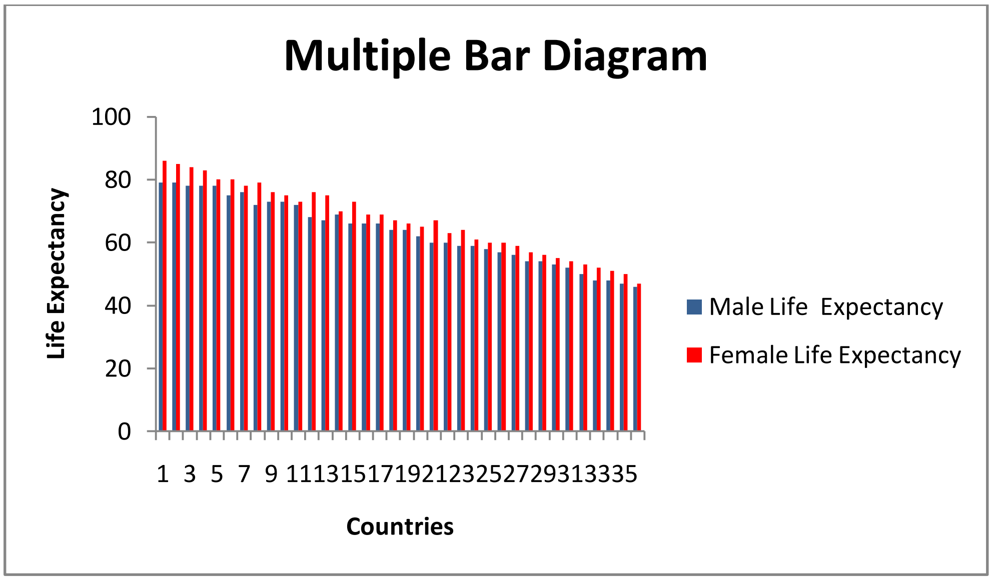




Interpretation:

- *It can be observed that the Female Life Expectancy of every country is greater than Male Life Expectancy hence <u>Women in a particular country live longer than Men</u>.*

MULTIPLE BAR DIAGRAM

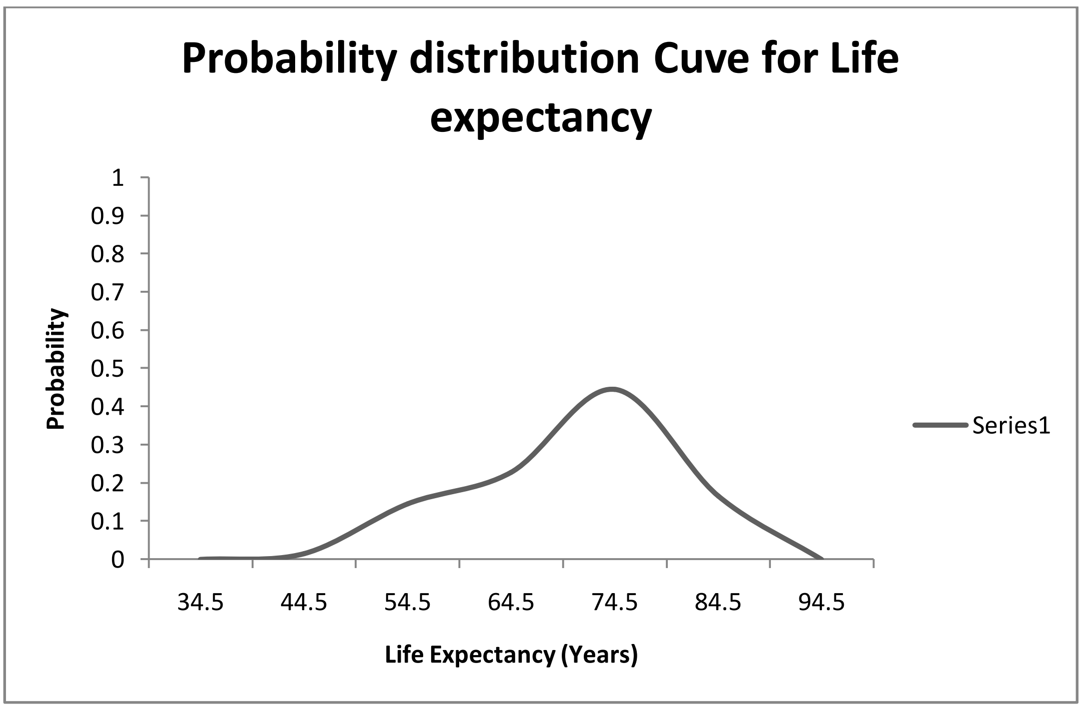


Interpretation:

- *It can be observed that the Female Life Expectancy of every country is greater than Male Life Expectancy hence* *__Women in a particular country live longer than Men__*.

COVERSION OF FREQUENCY DISTRIBUTION TABLE TO PROBABILITY DISTRIBUTION TABLE				
UPPER LIMITS	LABELS	FREQUENCY	CLASS MARK	PROBABILITY
49	40-49	3	44.5	0.015544041
59	50-59	28	54.5	0.14507772
69	60-69	44	64.5	0.227979275
79	70-79	86	74.5	0.445595855
89	80-89	32	84.5	0.165803109
		SUM=193		SUM=1

PROBABILITY DISTRIBUTION CURVE



Skewness	-0.690031347
Kurtosis	-0.444733331

Interpretation:

- *The Probability Distribution Curve of Life Expectancy is* <u>*negatively skewed*</u> *and* <u>*platykurtic*</u>*.*

COMPUTATION OF EXPECTED VALUE OF LIFE EXPECTANCY

CLASS MARK(x)	PROBABILITY[p(x)]	x.p(x)
44.5	0.015544041	0.691709845
54.5	0.14507772	7.906735751
64.5	0.227979275	14.70466321
74.5	0.445595855	33.19689119
84.5	0.165803109	14.01036269
	E[X]=70.5103626943005	

Interpretation:

- *The Expected Value of Life Expectancy obtained is* 70.51.
- *The values of Median & Mode calculated previously are* 73 & 74 *respectively.*
- *As* 70.51<73<74 i.e. *Mean<Median<Mode therefore the Probability Distribution of Life Expectancy is negatively skewed*.

BOX PLOT PREPARATION

Box Plot is used to compare two or more data sets using Five Number Summary.

Male Life Expectancy:

Sr. No.	Item	value	Error
1	Minimum	46	20
2	First Quartile	62	
3	Second Quartile	70	
4	Third Quartile	74	
5	Maximum	83	9

Female Life Expectancy:

Sr. No.	Item	value	Error
1	Minimum	47	19
2	First Quartile	66	
3	Second Quartile	76	
4	Third Quartile	80	
5	Maximum	86	6

Overall Life Expectancy:

Sr. No.	Item	value	Error
1	Minimum	46	18
2	First Quartile	64	
3	Second Quartile	73	
4	Third Quartile	76	
5	Maximum	86	10

COMPARISON OF MALE AND FEMALE LIFE EXPECTANCY USING BOX PLOT

(FIVE NUMBER SUMMARY)

$$Q_3 - Q_2 < Q_2 - Q_1$$

Bowley's Coefficient of skewness S_B=-0.5

Interpretation:

- *All the above three Box Plots are negatively skewed as the difference between the third and the second quartiles is less than that between the second and the first quartiles.(S_b =-0.5)*
- *The Interquartile range of Female Life Expectancy is greater than that of Male Life Expectancy. Hence the Spread of Female Life Expectancy is more than that of Male Life Expectancy.*
- *The Maximum and Minimum values of Female Life Expectancy are greater than those of Male Life Expectancy.*

REGRESSION ANALYSIS

(FOR THE RELATION BETWEEN GDP PER CAPITA OF COUNTRY AND IT'S LIFE EXPECTANCY.)

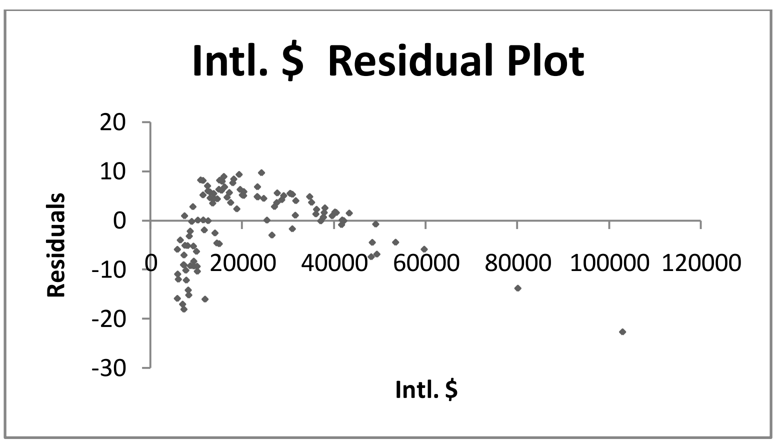
Let variable X: GDP per capita (PPP) of a country.

& variable Y: Life Expectancy of that Country.

GDP per capita: GDP per capita is often considered an indicator of a country's standard of living.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.671983
R Square	0.4515612
Adjusted R Square	**0.4485806**
Standard Error	6.7777317
Observations	186

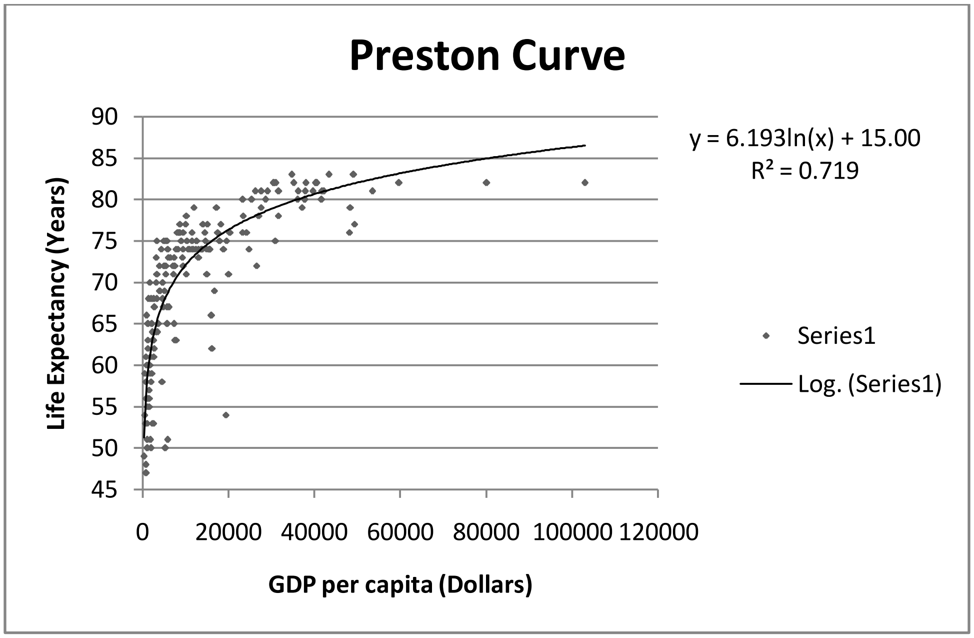


Interpretations:

- *There is a <u>Moderate type of Correlation</u> between the GDP per capita (PPP) of a country and its Life Expectancy.*
- *The Coefficient of determination (R Square) obtained is 0.4515612.Hence approximately 45% of variation in study variable is explained by Regression which is less.*
- *The Residual Plot obtained Shows a Curvilinear Pattern hence the Data needs to be Transformed.*

RESIDUAL OUTPUT (SAMPLE DATA)					
Observation	*Predicted Life Expectancy*	*Residuals*	*Observation*	*Predicted Life Expectancy*	*Residuals*
1	104.6706876	-22.67068761	21	79.15807146	-0.158071462
2	95.81993369	-13.81993369	22	78.80286245	2.197137555
3	87.90606295	-5.906062946	23	78.74624616	1.253753839
4	85.48629847	-4.486298468	24	78.38405705	3.615942945
5	83.90143029	-6.901430291	25	78.22273942	4.777260577
6	83.80564795	-0.805647947	26	77.04930876	3.950691236
7	83.51481087	-4.514810871	27	77.00781601	0.992183992
8	83.42600862	-7.426008617	28	76.7627373	5.237262702
9	81.56930472	1.430695281	29	76.75653277	-1.756532774
10	81.10900657	-0.109006572	30	76.62740111	5.372598888
11	80.96901699	0.030983007	31	76.5645803	5.435419697
12	80.91821745	-0.91821745	32	76.02556225	4.974437746
13	80.47226727	1.527732734	33	75.8580401	4.141959902
14	80.4152632	1.584736801	34	75.48033968	5.519660319
15	80.35321796	1.646782044	35	75.43690801	3.563091989
16	80.12248721	0.877512792	36	75.24534332	2.754656677
17	79.510566	2.489434001	37	75.03477728	-3.03477728
18	79.44696962	1.553030375	38	74.94752616	6.052473843
19	79.38492438	0.615075618	39	74.61170628	5.388293721
20	79.37833207	0.621667926	40	74.33948277	-0.339482775

RELATION BETWEEN GDP PER CAPITA (PPP) & LIFE EXPECTANCY FOR DIFFERENT COUNTRIES



Correlation Coefficient(r) = 0.671983

Interpretation:

- *There exists a Moderate type of Correlation between GDP per capita (PPP) and Life Expectancy of a country.*
- *It can be seen from the above graph that initially Life Expectancy increases with an increase in GDP per capita but at a certain point of GDP per capita it becomes constant & becomes invariant.*
- *The Graph attains stability as the man is immortal.*
- *Despite the U.S. having the highest health expenditure per capita, life expectancy in the US trails that in most other developed countries.*

Data Sources:

1) List of Countries by Life Expectancy: **WHO (Year 2011).**
2) List of Countries by GDP Per Capita: **CIA World Fact book 2011.**
3) Information of Continents**: Encyclopedia Britannica.**

Software Used:

1) **Microsoft Office Excel.**
2) **The "R"-Software for Statistics.**
3) **Microsoft Office Word.**

THE END