Statistics Investigation

Aim

The aim of investigation is to see if there is a link between the Gross National Product (GNP), average BMI (Body Mass Index) and average Life Expectancy of a country. I have chosen to do this because it gives a good insight into how healthy people in a country are and how good their standard of living is.

I have chosen the above factors to investigate because:

- GNP measures the value of goods and services that the country's citizens produced regardless of their location. GNP is one measure of the economic condition of a country, under the assumption that a higher GNP leads to a higher quality of living, all other things being equal. It is the best indicator of a countries wealth.
- BMI is a measure of body fat that is the ratio of the weight of the body in kilograms to the square of its height in meters. BMI is a better measure of health risk than actual weight in pounds. The medical terms, overweight and obesity, are based on BMI values. A BMI of between 25 and 30 is defined as overweight, and a BMI of 30 or more is considered obese. The higher your BMI, the greater the risk of developing a weight-related illness. It is best indicator of health as it shows average nutrition of a country.
- Life expectancy is the average lifespan of a person taken at birth. It is helpful because it shows standard pf living and available health resources in a country. It is also easy to compare with out country.

Hypotheses

From my investigation I think that:

- The longer the life expectancy, the larger the average BMI
- The larger the average BMI, the higher the GNP
- The higher the GNP, the longer the average life expectancy

I think this because the higher standard of living associated with More Economically Developed countries usually means that the population have better health care, have more national services and have better health care thus leading to a longer life expectancy, higher BMI and higher GNP.

How I am going to carry out the Investigation

A) The data needed for this investigation is the data from 30 different countries. I need to collect each countries average BMI, GNP and life expectancy. To make sure that this investigation is fair, I need to take results from MEDCs (more economically developed countries) and LEDCs (less economically developed countries) this will improve my investigation because it will produce a wider spread of results. I have chosen to investigate 30 countries as this will mean that I can do an even number of MEDCs and LEDCs.

B) The method of collect of my investigation will be secondary data from the Internet. I am using secondary data as it is impossible to send questionnaires out to 30 different countries. Also, the questionnaires might not go to a range of

people, which would lead to an unfair result. By using secondary data, I can make sure that I get a fair range of data from the website.

C) My sources of secondary data will be the World Health Organisation and the Royal Society of Medicine. I have chosen to find my data from these sources, as I believe that they will produce the most accurate results.

D) I am going to choose my data by selection. I will take 6 countries from each of the 5 World areas (such as Europe). To make my selections I am going to look at the average GNP as this is the set of data with the widest range of results and can show me which countries are MED or LED. Within this collection I shall get a range of the different GNPs for example I will choose countries with high GNPs, average GNPs and low GNPs. This, I feel, will give me the most accurate results as I will be using a range of data and no particular world area will have advantage over the others by only choosing countries with higher GNPs.

My Data

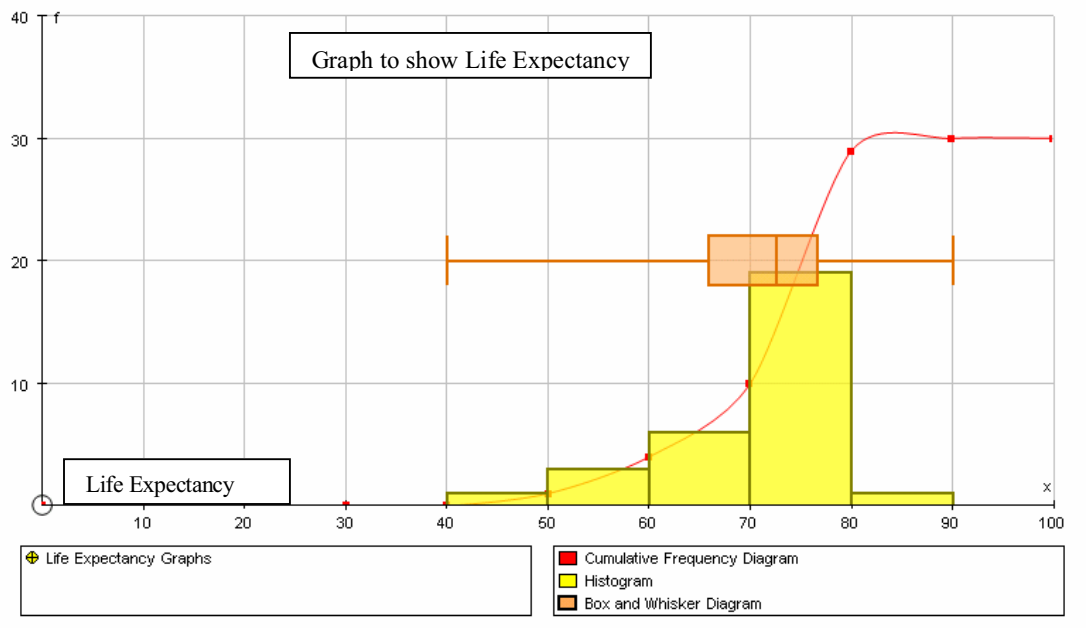
Country	Av. Life Expecta	Av. BMI	Av GNP
Croatia	73	26.17	4620
France	78	24.61	24210
Germany	77	26.56	26570
Italy	78	25.37	20090
Spain	78	25.79	14100
Sweden	79	25.75	25580
Switzerland	79	25.39	39980
UK	77	26.42	21410
Egypt	67	26.67	1290
Israel	78	25.79	16180
Jordan	71	26.06	1150
Morocco	67	23.24	1240
Tunisia	72	24.36	2060
Canada	79	26.72	19170
Haiti	54	21.31	410
Jamaica	75	23.81	1740
Mexico	72	26.61	3840
USA	77	27.83	29240
Brazil	67	24.48	4630
Ecuador	70	24.19	1520
Afghanistan	46	21.02	1
Australia	79	26.98	20640
Banglasdesh	59	20.05	350
Cambodia	54	21.69	260
India	63	21.37	440
Japan	81	22.96	32350
Pakistan	62	21.73	470
Singapore	77	22.56	30170
Thailand	72	23	2160
Vietnam	68	20.44	350

My data shows the country, average BMI, average life expectancy and average GNP. When looking for data on BMI, I could not find any on the Royal Society of Medicine's website so I contacted their Search services. With this investigation is a copy of the email that I sent to them and their reply from Emma Shaw, the Royal Society of Medicine's Search Assistant.

Analysis of the Data

- **Graph 1: Average Life Expectancy**

To show my results for Average Life Expectancy, (one of my three factors that I am researching) I have used the statistical computer programme, Autograph 3. To clearly show my results I have used a cumulative frequency diagram, a histogram and a box and whisker diagram. I have used a cumulative frequency graph to show the trend of growth of my continuous data. It is useful for estimating how much more or less there is of a certain amount, and keeps a running total of the amount of values. Histograms are summary graphs that show a count of data points falling in various ranges. The effect of this is a rough approximation of the frequency distribution of the data. I have used box and whisker diagrams because they are useful for showing median and upper and lower quartiles. They are also useful in seeing any outliners or anomalous results. I have also used stem and leaf diagrams to show my initial data as it is a clear was of representing data, where results are very easy to pick out and read off.



- Stem and Leaf Diagram for Life Expectancy Graphs:

40: 6
50: 4, 4, 9
60:2, 3,7,7,7,8
70:0,1,2,2,2,3,5,7,7,7,7,8,8,8,8,9,9,9,9
80: 1

- Statistics for Life Expectancy Graphs

Raw Data Statistics:

Number in sample, n:	30
Mean, x:	$\sum n/n = 2120/30 = 70.93$
Standard Deviation, x:	$\sqrt{\sum(x-\bar{x})2/n} = 8.6966$
Range, x:	81-46= 35
Lower Quartile:	$\frac{1}{4}$ (n+1)th value= 7.5^{th} value = 67
Median:	$\frac{1}{2}$ (n+1)th Value= 72.5
Upper Quartile:	$\frac{3}{4}$ (n+1)th value= 78
Semi I.Q. Range:	72.5-67= 5.5

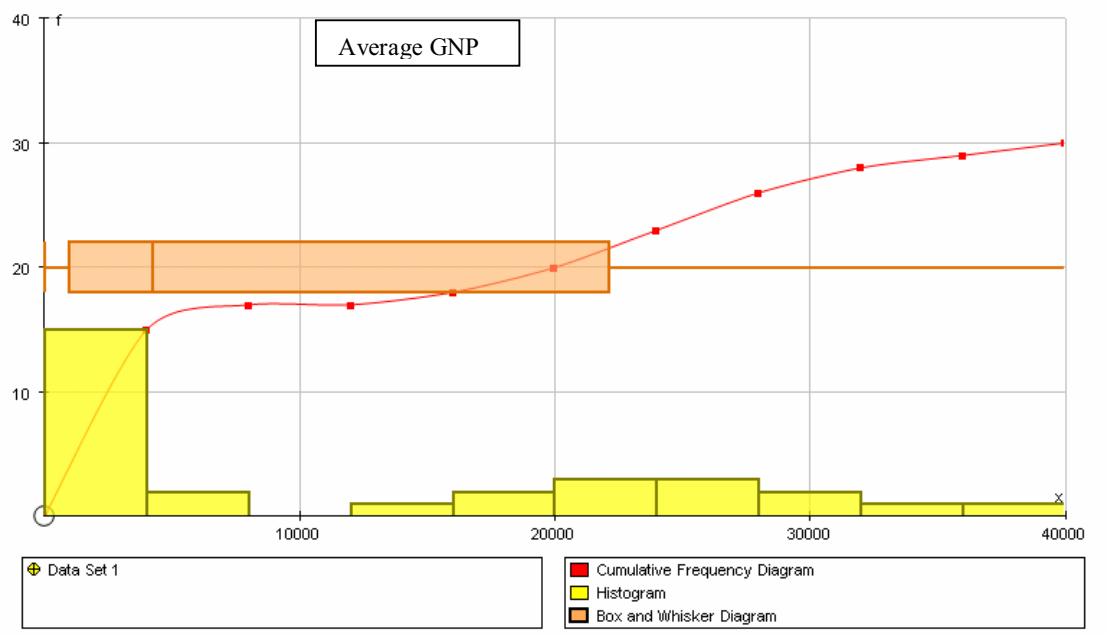
- Table of Values of Life Expectancy Graphs:

Class Int.	Mid. Int. (x)	Class Width	Freq.	Cum. Freq.
0 § x < 10	5	10	0	0
10 § x < 20	15	10	0	0
20 § x < 30	25	10	0	0
30 § x < 40	35	10	0	0
40 § x < 50	45	10	1	1
50 § x < 60	55	10	3	4
60 § x < 70	65	10	6	10
70 § x < 80	75	10	19	29
80 § x < 90	85	10	1	30
90 § x < 100	95	10	0	30

$\sum f = 30$
$\sum fx = 2110$
$\sum fx^2 = 1.506E+005$
Mean = 70.33
Standard Deviation = 8.459
Variance = 71.56

Analysis: From my data I can see that the average life expectancy is 70.9 years. The range of my results is 35, which suggests that there is a lot of difference between the life expectancy of MEDCs and LEDCs. The box and whisker diagram has shown me that the lower quartile 67years and the upper quartile is 78 years. This shows the measure of spread. Half of the values is 72.5years, $\frac{3}{4}$ is 78years and $\frac{1}{4}$ is 67 years.
The standard deviation is the square route of the variance, which measures the spread of the data. The spread of the data in this graph is 8.69. The advantage of using standard deviation is that it includes all of the data.

- **Graph 2: Average GNP**



- Results for Average GNP

Table of Values of Data Set 1:

Class Int.	Mid. Int. (x)	Class Width	Freq.	Cum. Freq.	
0 § x < 4000	2000	4000	15	15	
4000 § x < 8000		6000	4000	2	17
8000 § x < 12000		1E+004	4000	0	17
12000 § x < 16000		1.4E+004	4000	1	18
16000 § x < 20000		1.8E+004	4000	2	20
20000 § x < 24000		2.2E+004	4000	3	23
24000 § x < 28000		2.6E+004	4000	3	26
28000 § x < 32000		3E+004	4000	2	28
32000 § x < 36000		3.4E+004	4000	1	29
36000 § x < 40000		3.8E+004	4000	1	30

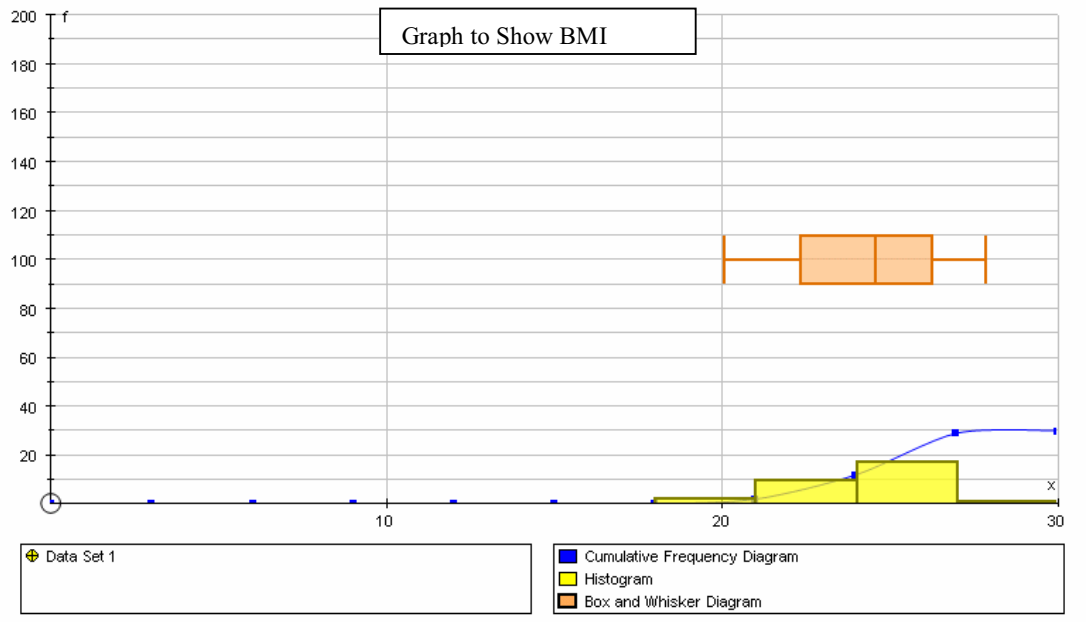
$\sum f = 30$
$\sum fx = 3.68E+005$
$\sum fx^2 = 8.856E+009$
Mean = 1.227E+004
Standard Deviation = 1.203E+004
Variance = 1.447E+008

- Raw Data Statistics:

Number in sample, n:	30
Mean, x:	11540.7
Standard Deviation, x:	12321
Range, x:	39979
Lower Quartile:	980
Median:	4230
Upper Quartile:	22110
Semi I.Q. Range:	10565

Again, my results for average Gross National Product have a huge range. This implies that there is a large difference in the amount spent on services and the value of goods for MEDCs and LEDCs. The mean shows that there are a large number of countries below the median whilst a few very rich MEDCs have the best services and the largest value of good. This indicates a large different between the very poor and the very rich.
This also shows that the countries with larger GNPs have much more money to spend on services for their populations and that their economies and produce of goods are strong, which often indicates a wealthy country.

- **Graph 3: Average BMI (Body Mass Index)**



- Stem and Leaf Diagram for Data Set 1:

20: 0.05, 0.44,
21: 0.02, 0.31, 0.37, 0.69. 0.73
22: 0.56, 0.96
23: 0.0, 0.24, 0.81,

24: 0.19, 0.36 0.48 0.61
25: 0.37, 0.39, 0.75, 0.79, 0.79
26: 0.06, 0.17, 0.42, 0.56, 0.61, 0.67, 0.72, 0.98
27: 0.83

- Statistics for Data Set 1

Raw Data Statistics:

Number in sample, n:	30
Mean, x:	24.2977
Standard Deviation, x:	2.19247
Range, x:	7.78
Lower Quartile:	22.3525
Median:	24.545
Upper Quartile:	26.2325
Semi I.Q. Range:	1.94

- Table of Values of Data Set 1:

Class Int.	Mid. Int. (x)	Class Width	Freq.	Cum. Freq.
0 § x < 3	1.5	3	0	0
3 § x < 6	4.5	3	0	0
6 § x < 9	7.5	3	0	0
9 § x < 12	10.5	3	0	0
12 § x < 15	13.5	3	0	0
15 § x < 18	16.5	3	0	0
18 § x < 21	19.5	3	2	2
21 § x < 24	22.5	3	10	12
24 § x < 27	25.5	3	17	29
27 § x < 30	28.5	3	1	30

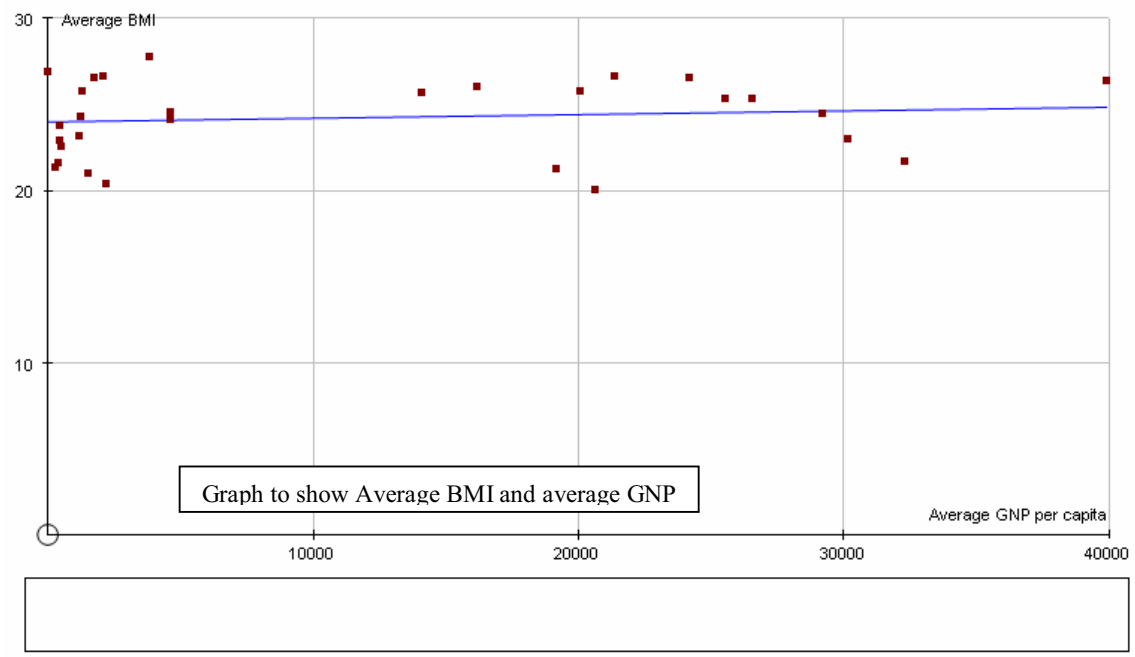
$\sum f = 30$
$\sum fx = 726$
$\sum fx^2 = 1.769E+004$
Mean = 24.2
Standard Deviation = 2.002
Variance = 4.01

- Analysis

From my results I can see that some countries have much higher BMI's than others. This indicates that some countries have much better standards of living and are able to have more food. The range of BMI is also significant as it demonstrates that food resources in the world are not shared out equally.
The average BMI is 24.2 which is about the average BMI expected for any population (average is 25, scource:WHO). However some results show that some countries have very low BMI, which indicates that the population in that country are undernourished and

unfed. However, some countries have a very high BMI, which indicates a problem with obesity and health.

- **Graph 4: Average GNP plotted against Average BMI**



- Results

For my results I am using Spearman's Rank Correlation Coefficient. This is a measure of the agreement between two sets of data. It is used to find the extent to which two sets of data correlate and is measured on a scale of –1 to +1 with one being a perfect positive agreement between the two sets on data, -1 being a perfect negative disagreement and 0 meaning now correlation. I am ranking each counties BMI and average GNP scores by the highest amount first. (B denotes average BMI, G denotes average GNP, a C denotes country)

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	Al	B2	C2	D2
	8	15	6	14	10.5	12	13	7	4	10.5	9	20	17	3	27	19	5	1	16	18	28	2	30	25	26	22	24	23	21	29
	15	7	5	10	13	6	1	8	21	12	23	22	18	11	26	19	16	4	14	20	30	8	27.5	29	25	2	24	3	17	27.5
	7	8	1	4	2.5	6	12	1	17	1.5	14	2	1	8	1	0	11	3	2	2	2	6	2.5	4	1	20	0	20	4	1.5
	49	64	1	16	6.25	36	144	1	289	2.25	196	4	1	64	1	0	121	9	4	4	4	36	6.25	16	0	400	0	400	16	2.25

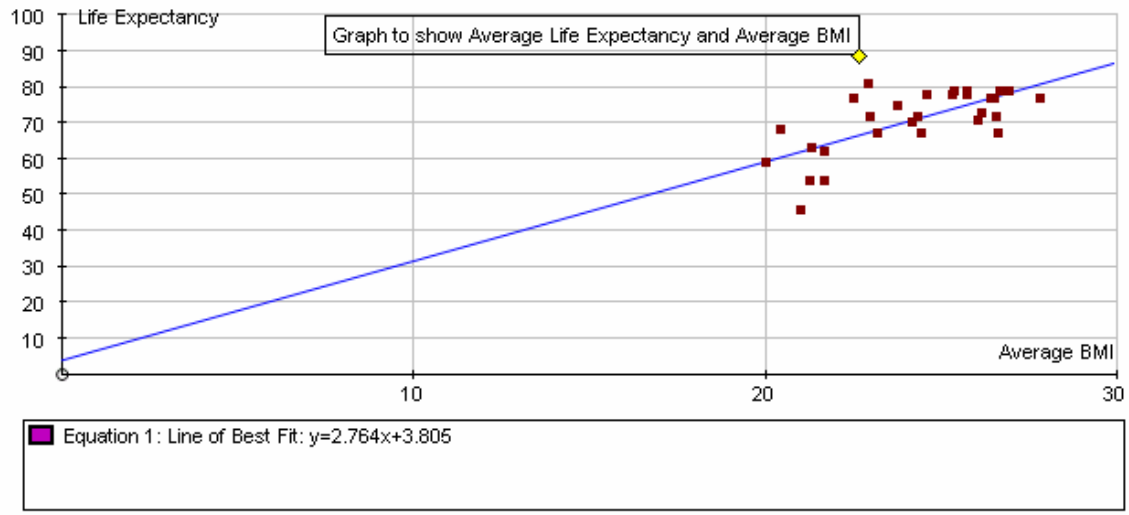
Spearman's coefficient= $1 - \frac{6\sum d2}{N(n2-1)}$

= 1- 6X1893/30(302-1)= 1- 11358/26970= 1-0.4211345= 0.58

- Analysis

From my spearman's rank results I can see that there is a good positive correlation between average BMI of a country and average GNP. I would expect this because a country with a higher average GNP would have more money to spend on food and going out therefore having a larger BMI. The countries with lower GNPs have much lower BMIs. Examples of these are the USA with a high average BMI of 27.83 and a high GNP of $29240 and Cambodia with a low GNP of 21.69 and a low GNP of $260. However, there are some anomalous results in my work for example Japan, who have the highest GNP but not the highest BMI. This may be because the Japanese eat healthy and culture in Japan promotes a healthy way of life. My results from this graph support my hypotheses.

- **Graph 5: Average Life Expectancy Plotted against Average BMI**



- Results

Again, I am using spearman's rank coefficient to see if these graphs have a relationship with each other. (b denotes BMI and l denotes life expectancy)

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	A1	B2	C2	D2
	8	15	6	14	10.5	12	13	7	4	10.5	9	20	17	3	27	19	5	1	16	18	28	2	30	25	26	22	24	23	21	29
	15	7.5	11.5	7.5	7.5	3.5	3.5	11.5	23	7.5	19	23	17	3.5	28.5	14	17	11.5	23	20	30	3.5	27	28.5	25	1	26	11.5	17	21

	7	7.5	5.5	6.5	3	8.5	9.5	4.5	19	3	10	3	0	0.5	1.5	5	12	10.5	7	2	2	1.5	3	3.5	1	21	2	11.5	4	8
	49	56.25	30.25	42.25	9	72.25	90.25	20.25	361	9	100	9	0	0.25	2.25	25	144	110.25	49	4	4	2.25	9	12.25	1	441	4	132.25	16	64

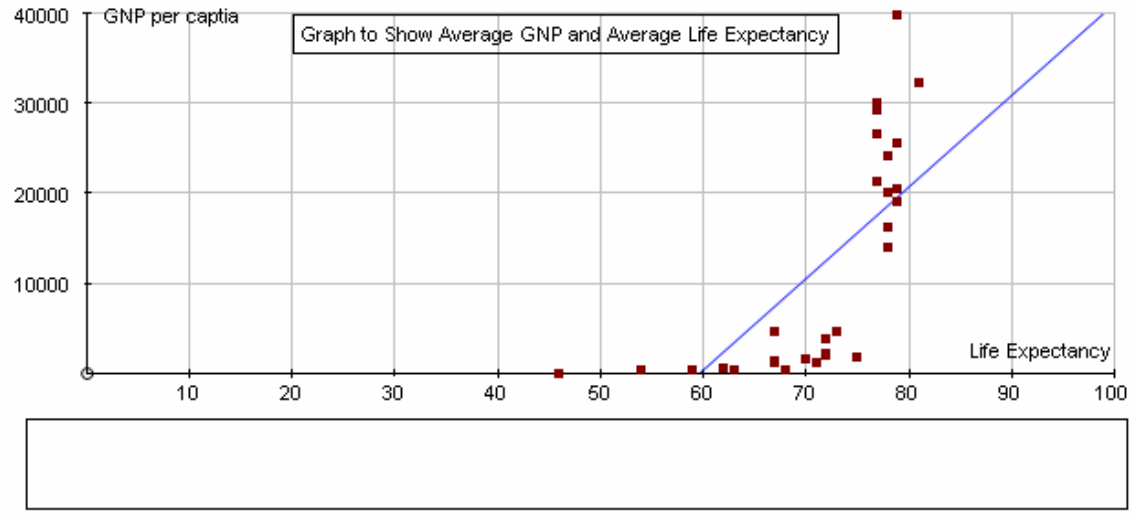
=Spearman's coefficient= $\frac{1- 6\sum d2}{N(n2 -1)}$

= 1- 6x1869/30(302-1)= 1- 11214/26970= 1-0.4157=0.584

- Analysis

From my results I can see that there is a positive link between average BMI and average GNP. This supports my hypothesis, but only to a certain point. I would expect this as countries with an average or high BMI obviously have access to better and more food than those with a BMI and amount and quality of food eaten is a deciding point of life expectancy as nutrition leads to an early death. An example of this is Switzerland with a slightly above average BMI and the second highest average Life Expectancy. However, in my hypothesis I stated that 'The longer the life expectancy, the larger the average BMI', this is only true to a certain extent because obese BMIs obviously cut down life span as when you are obese you are more likely to have heart problems and diabetes. The countries with the highest average life expectancy were those with the average BMI of 25, or slightly more or less.

- **Graph 6: Average Life Expectancy plotted against Average GNP**



- Results

I have used Spearman's Rank Coefficient to show the relationship between average Life Expectancy and Average GNP. (l denotes life expectancy and g denotes average GNP)

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	A1	B2	C2	D2
	15	7	5	10	13	6	1	8	21	12	23	22	18	11	26	19	16	4	14	20	30	8	27.5	29	25	2	24	3	17	27.5
	15	7.5	11.5	7.5	7.5	3.5	3.5	11.5	23	7.5	19	23	17	3.5	28.5	14	17	11.5	23	20	30	3.5	27	28.5	25	1	26	11.5	17	21
	0	0.5	6.5	2.5	5.5	2.5	2.5	3.5	2	4.5	4	1	1	7.5	2.5	5	1	7.5	9	0	0	4.5	0.5	0.5	0	1	2	8.5	0	6.5
	0	0.25	42,25	6.25	30.25	6.25	6.25	12.25	4	20.25	16	1	1	56.25	6.25	25	1	56.25	81	0	0	20.25	0.25	0.25	0	1	4	72.25	0	42.25

=Spearman's coefficient= $\frac{1- 6\sum d2}{N(n2 -1)}$

= 1- 6X512/30(302-1)= 1-3072/26970= 1-0.114=0.886

- Analysis

From my results, I can see that there is a very strong link between Average GNP and Average Life Expectancy. This supports my hypothesis and I would expect this because a higher GNP means that the country has more money to spend on services such as hospitals and doctors for its population, which leads to better health care and a longer life expectancy. A good example of this is Switzerland with the highest average GNP and one of the longest life expectancies. The only anomalous result that I had was Japan who the largest Life Expectancy which suggests that the health care and standard of living in Japan is very good.

N.B Point on Standard Deviation when calculated some of the calculations showed xE+E or for example 5.06E+05. This is the calculator's way of doing and simplifying Standard Form, 5.06E+05 means 5.06 x 10 ^ (+05) or 5.06 x 10^5 or five hundred and six thousand 506,000

Summary and Conclusion of Data

What does my data show: I have found this investigation very interesting and it has been interesting to look at different sets of data plotted against each other. I believe that my data has shown my hypothesis with the MDEDS having much longer life expectancies, BMIs and GNPs than LEDCs. My results have proven that, on average, countries with higher GNPs live longer and have higher BMIs. I reasons for my results are that MEDCs have better standards of living with food readily available and good health services. MEDCs also have good with better values, which improves wages and standard of living.

Limitations of my work: I felt that within my investigation there were several limitations: The size of the populations may have made the investigation unfair as a larger population would have brought the average life expectancy up as there was more people to measure the ages of.

The data that I used was collected in 1998. This was several years ago and average BMI's, Life expectancies and average GNPs would have changed, meaning that my data and research is not valid, as new data has been found.
Also, the World Health Orgainsation only had all of my relevant data for certain countries so I did not have many countries to choose from which may have limited my results.
I also felt that I would have gotten better results if I had more time to investigate more countries, as it would have made a better comparison. However, I am pleased with the amount of countries that I have investigated in this investigation.

Points for further work:
I have two possible sets of inter- related hypothesis that I could suggest for further work.

1. Does the geographically position of the country affect the BMI of the population? Does the geographically position affect the GNP of a country? Does the average GNP affect the infant mortality rate? Does the average BMI affect the average infant morality rate?
2. Does the average life expectancy affect the average number of children per household? Does the average GNP affect the number of children per household? Does the average BMI affect the number of children per household?

Sources of Information
I collected my data and information from the following places;

- The World Health Orgainsation www.who.int
- Royal Society of Medicine www.rsm.ac.uk
- The World Bank www.worldbank.org
- GCSE Geography Dictionary
- Statistics GCSE for AQA