Data Analysis Application Task: HAPPINESS

<u>**Part 1**</u>**: Is there a trend in Happiness from 2013 and 2021? (29 marks)**

COUNTRY	HAPPINESS 2013	HAPPINESS 2021
Australia	7.4	6.5
Afghanistan	7.4	7.2
Azerbaijan	4.6	5.2
Cambodia	4.1	4.6
Croatia	4.1	4.3
China	5.0	5.6
Denmark	7.7	7.6
Finland	7.4	7.8
France	6.8	6.7
Hungary	4.8	6.1
Iceland	7.4	7.6
Indonesia	5.3	5.2
Ireland	7.1	7.0
Italy	6.0	6.5
Japan	6.1	6.0
Luxembourg	7.1	7.4
Malta	6.0	6.5
Mauritius	5.5	6.1
Moldova	5.8	5.9
New Zealand	7.2	7.2
Nicaragua	5.5	6.2
Norway	7.7	7.4
Philippines	5.0	5.9
Romania	5.0	6.5
Singapore	6.5	6.5
South Korea	6.3	5.9
Sudan	6.3	7.2

For Happiness Index 2013:

Question 1: Find the mean, *correct to 1 decimal place.*

Question 2: Find the median, *correct to 1 decimal place.*

Question 3: Find the Lower Quartile, *correct to 1 decimal place.*

Question 4: Find the Upper Quartile, *correct to 1 decimal place.*

Question 5: Find the Interquartile Range, *correct to 1 decimal place.*

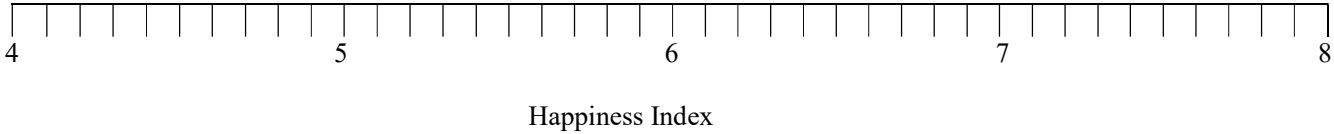**Question 6:** Find the Upper Fence, *correct to 1 decimal place.*

Question 7: Find the Lower Fence, *correct to 1 decimal place.*

Question 8: Are there any outliers? Why/Why not?

Question 9: Calculate the summary statistics and IQR for the **2021** Happiness Index correct to 1 decimal place.

Question 10:   For the Happiness Index data of 2021 are there any outliers? (Conduct an outlier test for the **2021** Happiness Index).

_____

_____

_____

_____

_____

_____

Question 11:   Construct parallel boxplots on the axis below for the 2013 and 2021 Happiness Indices. Label each boxplot clearly.

```
 |                         |                        |                        |                        |
 4                         5                        6                        7                        8
```

Happiness Index

Question 12:   Describe the shape of each distribution (2013 and 2021) from the Boxplot.

_____

_____

_____

_____

_____

_____

Question 13:   Compare the Happiness Index from 2013 and 2021 by quoting an appropriate measure of centre.

_____

_____

_____

_____

_____

_____

Question 14:   Compare Happiness Index from 2013 and 2021 by quoting a statistic other than a measure of centre.

_____

_____

_____

_____

_____

_____

Question 15: Which summary statistic, mean or median, would be most appropriate to describe the central tendency of each data set (2013 and 2021). Explain.

_____

_____

_____

_____

_____

_____

_____

COUNTRY		HAPPINESS 2021	LIFE EXPECTANCY 2021
Australia		6.5	82.3
Afghanistan		7.2	81.2
Azerbaijan		5.2	66.9
Cambodia		4.6	60.4
Croatia		4.3	52.8
China		5.6	78.1
Denmark		7.6	81.6
Finland		7.8	82.1
France		6.7	82.2
Hungary		6.1	75.6
Iceland		7.6	83.1
Indonesia		5.2	68.8
Ireland		7.0	82.2
Italy		6.5	82.3
Japan		6.0	84.6
Luxembourg		7.4	81.7
Malta		6.5	82.7
Mauritius		6.1	74.2
Moldova		5.9	70.2
New Zealand		7.2	82.1
Nicaragua		6.2	71.8
Norway		7.4	83.2
Philippines		5.9	72.1
Romania		6.5	74.4
Singapore		6.5	83.7
South Korea		5.9	83.4
Sudan		7.2	83.2

Below is a scatterplot showing the relationship between Happiness Index and Life Expectancy in 2021 for the 27 countries listed in the data sample.
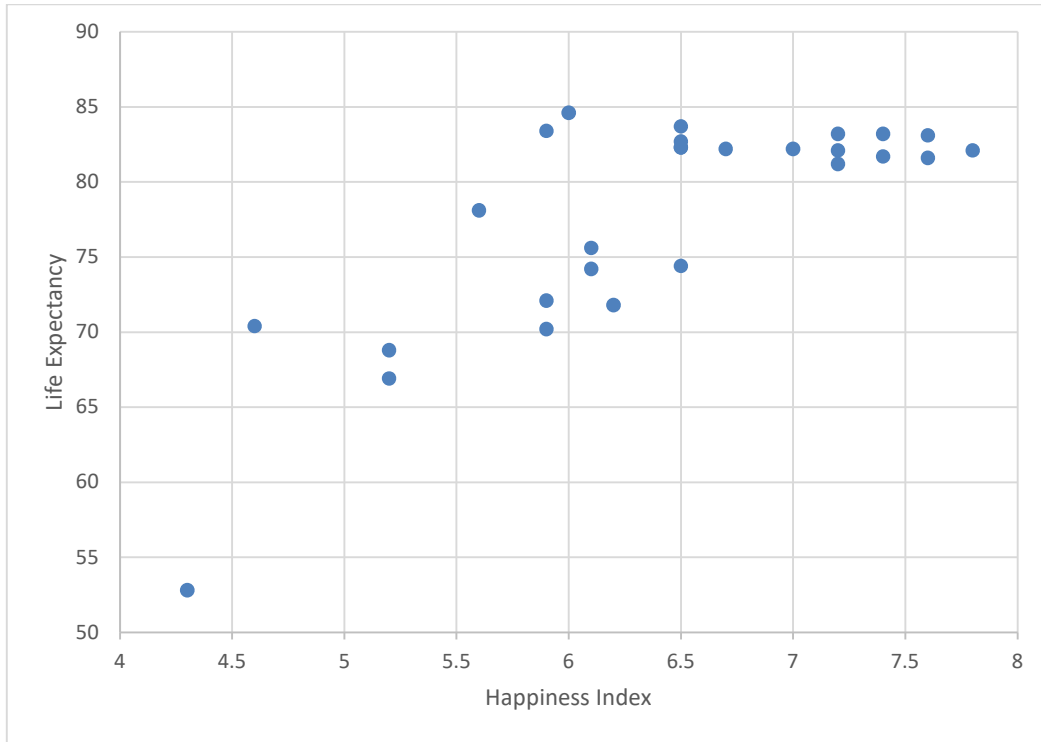
Question 16:   Write down the explanatory and response variables.

_____

_____

_____

_____

Question 17:   Determine the equation of the least squares regression line for the scatterplot above. Write the equation in terms of the variables in the form $y = a + bx$ and state the coefficients correct to 2 significant figures.

_____

_____

_____

_____

Question 18:   Draw the least squares regression line found in Question 17 on the scatterplot above. State the two points used to draw this line.

Question 19:   Interpret the slope of this least squares regression line in terms of the variables 2021 Happiness Index and 2021 Life Expectancy.

Question 20:   Interpret the vertical intercept of the least squares regression line in terms of the variables 2021 Happiness Index and 2021 Life Expectancy.

Question 21: In 2021, the Happiness Index for Burundi was 2.9. Calculate the predicted Life Expectancy for Burundi in 2021 (as a whole number) and comment on the reliability of your answer.

_____

_____

_____

_____

Question 22: In 2021 the Happiness Index for France was 6.7 and Life Expectancy was 82.2. Calculate the residual when the least squares regression line is used to predict the 2021 Life Expectancy from 2021 Happiness Index. Round the answer to 1 decimal place.

_____

_____

_____

_____

_____

_____

Question 23: Calculate the correlation coefficient for the 27 countries in 2021 correct to 3 decimal places and comment on what this indicates about the relationship between Happiness Index and Life Expectancy in 2021.

_____

_____

_____

_____

_____

_____

Question 24: Calculate the coefficient of determination for the 27 countries in 2021 correct to 2 decimal places and comment on what this indicates about the relationship between Happiness Index and Life Expectancy in 2021

_____

_____

_____

_____

_____

_____

_____

Part 2B: Transforming the data (17 marks)

Refer to the scatterplot in Part 2A on page 8.
In an attempt to improve the linearity of the data, we will look at different transformations.

Question 25: Suggest three possible transformations:

_____

_____

_____

Question 26: By applying each of the 3 transformations to your data. Complete the table below for the transformed data values. Use y to represent 2021 Life Expectancy rate and x to represent 2021 Happiness Index in your equations.

Transformation	Equation of least squares regression line _(Round all values to 2 decimal places)_	Correlation Coefficient _(Round all values to 4 decimal places)_

Question 27: Comment on the improvement or otherwise of each transformation equation over the original equation. Use mathematical reasoning to support your comments. Choose the best equation for calculating predicted values of the four considered (original from Question 17 plus 3 transformations).

_____

_____

_____

_____

_____

_____

Question 28: Use your chosen equation (with previously rounded coefficients) to predict a country's Life Expectancy rate in 2021 if it's Happiness Index in 2021 was 9.5.
Round your answer to the nearest whole number.

_____

_____

_____

_____

_____

Question 29: Use your chosen equation to predict a country's Happiness Index in 2021 if it's Life Expectancy in 2021 was 75.
Round your answer to one decimal place.

_____

_____

_____

_____

_____

Question 30: Which prediction from Question 28 and Question 29 is more reliable. Explain. State whether it is interpolation or extrapolation.

_____

_____

_____

_____

_____

_____

Question 31: Construct a Stem and Leaf Plot of the 2021 Happiness Index from the following table:

Australia	Afghanistan	Azerbaijan	Cambodia	Croatia	China	Denmark	Finland	France
6.5	7.2	5.2	4.6	4.3	5.6	7.6	7.8	6.7
Hungary	Iceland	Indonesia	Ireland	Italy	Japan	Luxembourg	Malta	Mauritius
6.1	7.6	5.2	7	6.5	6	7.4	6.5	6.1
Moldova	New Zealand	Nicaragua	Norway	Philippines	Romania	Singapore	South Korea	Sudan
5.9	7.2	6.2	7.4	5.9	6.5	6.5	5.9	7.2

Question 32: Describe the shape of the Stem and Leaf Plot.

_____

_____

_____

_____

Question 33: What percentage of countries had a Happiness Index higher than 6.0 in 2021? State your answer to the nearest whole number.

_____

_____

_____

_____

Question 34: Why would it not be suitable to calculate the z-score for 2021?

_____

_____

_____

_____

_____

Part 3B: Normal Distribution and Time Series (17 marks)

In order to cater for visitors to their Happy country, a city in Finland is planning to build a new Hotel. The projections indicate that if a 500 room hotel is built, the mean occupancy will be 150 rooms with a standard deviation of 25. Assume that the occupancy is normally distributed.

Question 35: What percentage of the time will the occupancy exceed 175 rooms?

_____

_____

Question 36: Calculate the number of nights per year that the occupancy will be below 125 rooms. Round your answer to the nearest whole number.

_____

_____

_____

Question 37: The standardised number of rooms occupied, z, on a particular night is 1.8 Calculate how many rooms are actually occupied.

_____

_____

_____

The time series graph below shows the Happiness Index trend in Sweden from 2004-2021.

Sweden Happiness Index

Question 38: Use this time series graph to describe the trend of the Happiness Index in Sweden from 2004 to 2021.

_____

_____

_____

_____

Question 39: Smooth the graph above using a 3-point median smoothing graphical method. Mark in your points using a X.

Question 40: Use the table below to complete a 5-point moving mean. Round all values to 1 decimal place.

Year	Life Expectancy	5 point moving mean
2005	6.1	
2006	6.2	
2007	6.3	
2008	5.9	
2009	6.4	
2010	6.6	
2011	6.7	
2012	6.4	
2013	6.9	
2014	7	
2015	7.1	
2016	6.8	
2017	7.2	
2018	7.3	
2019	7.4	
2020	7.1	
2021	7.4	

Question 41: Plot the 5-point moving mean data onto the scatterplot above using a ●

Question 42: Comment on the effectiveness of the two smoothing techniques to identity any other trends.

_____

_____

_____

_____

The data for Happiness Index in Australia for 2017 to 2019 is presented in the table below.

	Summer	Autumn	Winter	Spring
2017	7.5	7.1	7.2	7.7
2018	7.4	7	7.2	7.6
2019	7.3	7	7.1	7.7

Question 43:   Plot the data as a time series plot using an appropriate time code.

Happiness Index

Question 44:   Calculate the yearly averages correct to 3 decimal places.

	Summer	Autumn	Winter	Spring	Yearly Average
2017	7.5	7.1	7.2	7.7	
2018	7.4	7	7.2	7.6	
2019	7.3	7	7.1	7.7	

Question 45:   Complete the table below by filling in the shaded boxes (correct to three decimal places)

	Summer	Autumn	Winter	Spring
2017	1.017			1.044
2018		0.959	0.986	
2019	1.003	0.962		
Seasonal Indices	1.011		0.979	

Question 46:   Deseasonalise the data using seasonal indices (correct to one decimal place)

	Summer	Autumn	Winter	Spring
2017				
2018				
2019				

Question 47:   Plot the original and deseasonalised data on the classpad **and** comment on the effectiveness of deseasonalising the data and identifying the trend.

_____

_____

_____

_____

_____

Question 48:   Determine the equation of least square regression line for the deseasonalised data, in the form: $y = a + bx$ correct to 3 significant figures.

_____

_____

_____

_____

_____

_____

_____

Question 49:   Use the equation above to predict the **actual** Happiness Index for Australia in winter 2023 to 3 decimal places.

_____

_____

_____

_____

_____

_____

END OF TASK