Chapter 4 Data transformation: **Assignment**

Student name: _____

The following tables gives the GDP (in $US per capita) and life expectancy (in years) for a group of 20 countries in 2020:

Country	*Life Expectancy*	*GDP ($US per capita)*
Afghanistan	65.17	516.75
Australia	83.20	51680.32
Belgium	80.80	45189.37
Brazil	76.08	6814.88
China	77.10	10408.67
Ecuador	77.22	5627.77
Greece	81.09	17647.23
India	69.89	1933.10
Iran	76.87	2756.75
New Zealand	82.06	41596.51
Papua New Guinea	64.73	2757.01
Philippines	71.36	3301.22
Singapore	83.74	60729.45
Slovak Republic	76.87	19266.51
Somalia	57.70	438.26
South Africa	64.38	5655.87
Spain	82.33	27056.42
Uganda	63.71	822.03
United Kingdom	80.90	41098.08
United States	77.28	63027.68

1

a We wish to predict *life expectancy* from *GDP*. Which would be the response variable, and which would be the explanatory variable?

b Construct a scatterplot of *life expectancy* against *GDP*, and describe the association in terms of direction, form and strength.

c Use the circle of transformations on page 239 to confirm that the transformations which might linearise the relationship between *life expectancy* and *GDP* are y^2, $\log x$ and $1/x$.

Chapter 4 Data transformation: **Assignment**

2 The y^2 transformation:

a Construct a scatterplot of (*life expectancy*)^2 against *GDP*.

b Fit least squares regression model which enables (*life expectancy*)^2 to be predicted from *GDP*. Write down the equation of the model giving the values of the coefficients rounded to five significant figures.

c Construct a residual plot, and comment on whether the linearity assumption has been met.

d Determine the value of the coefficient of determination for this model. Give your answer as a percentage rounded to one decimal place.

3 The log(x) transformation:

a Construct a scatterplot of *life expectancy* against log(*GDP*), and describe the association in terms of direction, form and strength.

b Fit least squares regression model which enables *life expectancy* to be predicted from log(*GDP*). Write down the equation of the model giving the values of the coefficients rounded to five significant figures.

c Construct a residual plot, and comment on whether the linearity assumption has been met.

d Determine the value of the coefficient of determination for this model. Give your answer as a percentage rounded to one decimal place.

4 The 1/x transformation:

a Construct a scatterplot of *life expectancy* against 1/*GDP*, and describe the association in terms of direction, form, and strength.

b Fit least squares regression model which enables *life expectancy* to be predicted from 1/*GDP*. Write down the equation of the model giving the values of the coefficients rounded to five significant figures.

c Construct a residual plot, and comment on whether the linearity assumption has been met.

d Determine the value of the coefficient of determination for this model. Give your answer as a percentage rounded to one decimal place.

5

a Based on the previous analyses, recommended a regression model for the association between *life expectancy* and *GDP*.

b Use your recommended model to predict, rounded to one decimal place, the *life expectancy* of a country where the *GDP* is*:*

i. $US20 000 per capita

ii. $US80 000 per capita

c Comment on the reliability of your predictions in part b.