Problem Set

STAT 101
UAEU

This is a Student's Activity Task containing a few multiple type questions and a number of descriptive type problems. The activity is not a graded component of the course. Its objective is to encourage students learning while solving problems and also to prepare for the upcoming exam.

Must Review: Review All the Quiz Problems.

1. Let us consider a simple linear regression betwen the response variable Y and the covariate X. The fitted regression line is $\hat{Y} = 3 + 2.5X$. Then idensity which if the following statement is correct?

   Ans:

   ☐ The correlation between X and Y must be positive.

   ☐ The correlation between X and Y must be negative.

   ☐ The correlation between X and Y must be zero.

   ☐ From the provided information, it is not possible to determine whether the correlation is positive, negative or zero.

2. **Statement:** If correlation between two continuous variables is zero then the variables must be unreleated.

   Ans: ☐ TURE   ☐ FALSE

3. **Statement:** If two continuous variables are unrelated then the correlation between them must be zero.
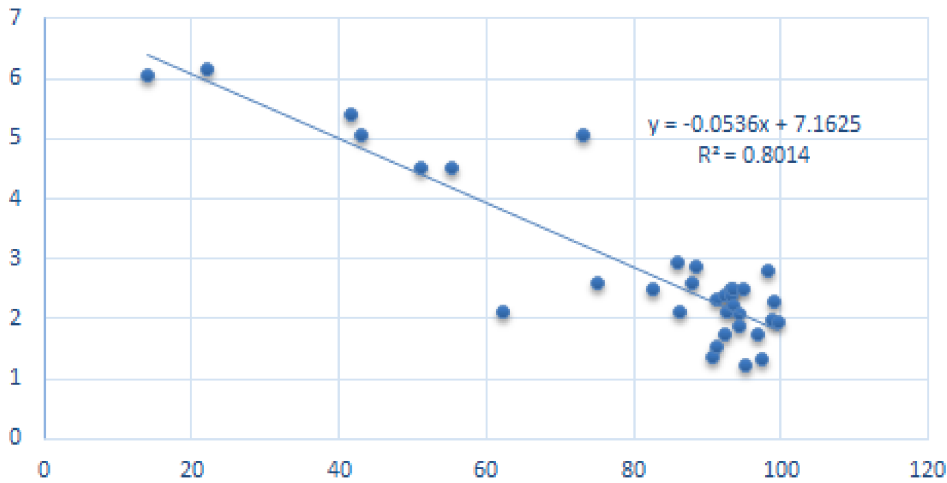
   Ans: ☐ TURE   ☐ FALSE

4. Consider the scatter plot and the corresponding regression line for the 'Adult Female Literacy' (x) and on 'Total Fertility Rate' (y) based on a sample of 35 countries worldwide.

TFR

$y = -0.0536x + 7.1625$
$R^2 = 0.8014$

Based on the information provided on the plot, what can we tell about the correlation between the two two variables?

Ans: 0.8952   -0.8952   0.8014   Can not be determined

5. If the Gini coefficient for Estonia income is 0.348 in 2015 and the Gini coefficient for Macedonia income is 0.434 in the same year, which country has the least unequal distribution of income among the two?

Ans: Macedonia   Estonia   Can be determined

6. Consider the Lorenz curves of wealth inequality for Japan and USA. Based on the plot below, which country has less wealth inequality among the two?
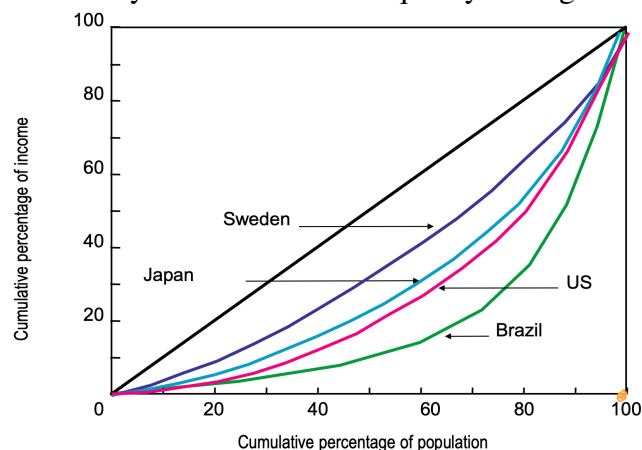
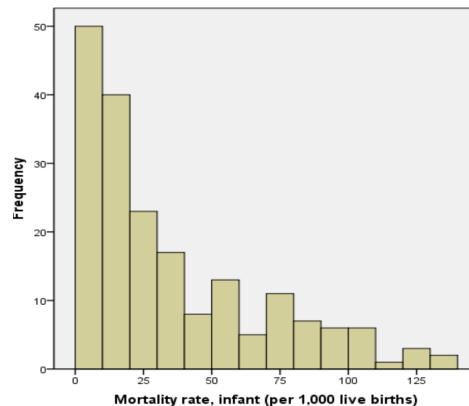Ans: Japan   USA   Can be determined

7.

Consider the Lorenz curves of wealth inequality for Brazil and Sweden. Based on the plot below, which country has less wealth inequality among the two?
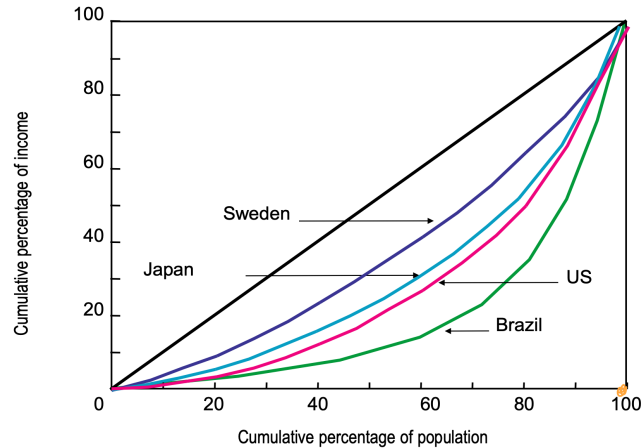
Ans: ☐ Brazil   ☐ Sweden   ☐ Can be determined

8.

Consider the histogram of 'Infant Mortality Rate'

Based on the provided histogram, what can we say about the type of the histogram of the variable? There are multiple corect answers, select all the correct options.

Ans:

☐ Skewed Right   ☐ Skewed Left   ☐ Skewed Positive

☐ Skewed Negative   ☐ Mean is larger than Median

☐ Mean is smaller than Median

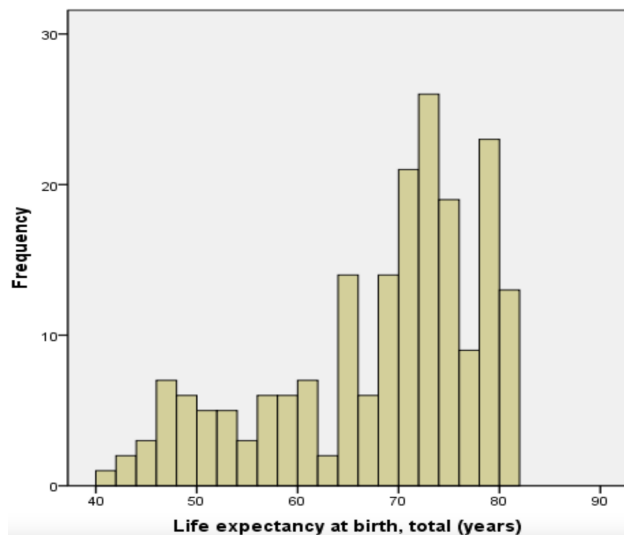**9.** Consider the histogram of 'Life Expectancy at birth'

Based on the provided histogram, what can we say about the type of the histogram of the variable? There are multiple corect answers, select all the correct options.

Ans:

☐ Skewed Right   ☐ Skewed Left   ☐ Skewed Positive

☐ Skewed Negative   ☐ Mean is larger than Median

☐ Mean is smaller than Median

10. According to the 2007 World Bank data, 5.21at $1.25 a day (PPP). When 5000 individuals were selected from Brazil, 6% were living below the $1.25 poverty line. What does the value 5.21% refer to?

Ans: ☐ Statistic   ☐ Parameter   ☐ Population   ☐ Sample

11. According to the 2007 World Bank data, 5.21at $1.25 a day (PPP). When 5000 individuals were selected from Brazil, 6% were living below the $1.25 poverty line. What does the the value 6% refer to?

Ans: ☐ Statistic   ☐ Parameter   ☐ Population   ☐ Sample

12. In the 2009, American Community Survey conducted to determine, among other things, the family size, 30366 American households were randomly sampled. The sample mean and the standard deviation of the number of persons in family were 3.11 and 1.723. Compute a 90% confidence the average family size in USA.

Ans:

According to the 2007 UNDP data, the mean and the standard deviation of the GDP index for 182 countries are 0.6925 and 0.2164 respectively. Compute a 99% confidence interval for the average GDP index over the last decade .

Ans:

13.

Consider the following summary of two continuous variables X, Y of a dataset.

Sample Mean of X	Sample Mean of Y	Sample Standard Deviation of X	Sample Standard Deviation of Y	Correlation between X, and Y
45	30	5	7.5	.9

Finally we consider a simple linear regression model: $\hat{Y} = a + bX$ where a and b denotes the intercept and the slope corrrespondingly. Estimate a, b and writedown the equation of the fitted line.

14.

Ans:

1. ## Must Review: Review All the Quiz Problems.

To Estimate the average life expectancy in **less-Developed countries**, a random sample of 144 less developed countries reveals the following data. Sample size:144, Mean:67.1 Standard deviation:28.9

On the other hand, the average life expectancy in Developed countries, a random sample of 144 **developed countries** reveals the following data. Sample size:144, Mean:76 Standard deviation:28.9

2.

(a) Compute the 99% confidence interval for the average life expectancy in less-developed countries.
(b) Compute the 99% confidence interval for the average life expectancy in developed countries.
(c) Explain whether there is a difference in the life expectancy between less developed and developed countries.

3.

Consider an example of the income distribution given by.

	Income %	Cumulative % Income
Income share held by lowest 20%	10	
Income share held by second 20%	15	
Income share held by third 20%	20	
Income share held by fourth 20%	25	
Income share held by highest 20%	30	

(a) Complete the table with the cumulative percentages of the income.
(b) Plot the Lorenz curve of this hypothetical country.

Ans:

a)

	Income %	Cumulative % Income
Income share held by lowest 20%	10	
Income share held by second 20%	15	
Income share held by third 20%	20	
Income share held by fourth 20%	25	
Income share held by highest 20%	30	

4.

Consider an example of the income distribution given by.

	Income %
Income share held by lowest 20%	10
Income share held by second 20%	15
Income share held by third 20%	20
Income share held by fourth 20%	25
Income share held by highest 20%	30

  (a) Complete the following table and compute the Gini's Index for Income Inequality

Ans

Decile	f_i	x_i	y_i	$y_i + y_{i-1}$	$f_i(y_i + y_{i-1})$
10	10	2.1	2.1	2.21	21
20	10	3.5	5.6	7.7	77
40	20	9.9	15.50	21.1	422
60	20	14.2	29.7	45.2	904
80	20	21.1	50.8	80.5	1610
90	10	15.7	66.5	117.3	1173
100	10	33.5	100	166.52	1665
				Total=	**5872**

$$G = 100\% - \left(\frac{5872}{10,000} \times 100 \right) = 41.28\%$$

5.

The following table is a hypothetical population of 100 individuals with a maximum life span of 4 years, using the remaining life expectancy method:

(a) Complete the following table.
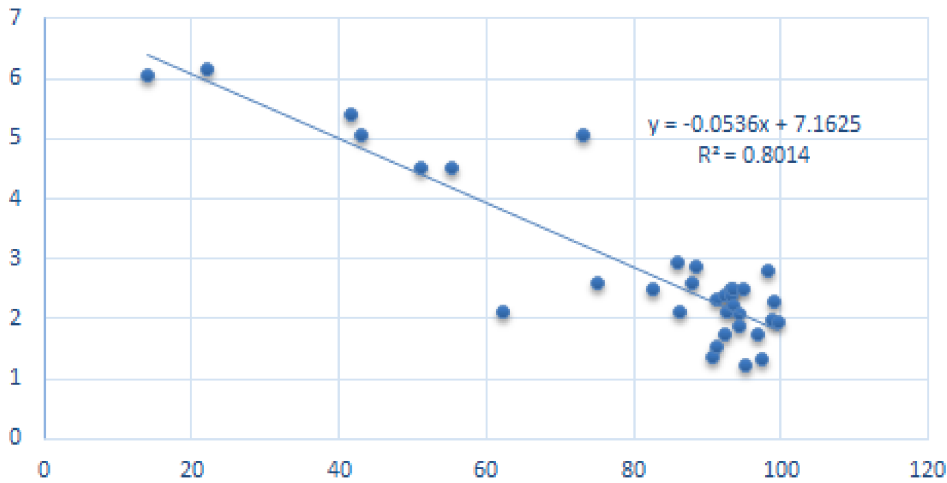
x	l_x	d_x	L_x	T_x	e_x
0	100	25			
1		10			
2		25			
3		40			
4		0			

(b) What is the life expectancy at birth of this hypothetical population?

Ans:

x	l_x	d_x	L_x	T_x	e_x
0	100	25			
1		10			
2		25			
3		40			
4		0			

6.

Consider the scatter plot and the corresponding regression line for the 'Adult Female Literacy' (x) and on 'Total Fertility Rate' (y) based on a sample of 35 countries worldwide.

TFR

$y = -0.0536x + 7.1625$
$R^2 = 0.8014$

Based on the information provided on the plot, answer the following questions:
- (a) Provide the value of the **slope** and its **interpretation**.
- (b) Provide the value of the intercept and explain the corresponding interpretation. Is the interpretation meaningful in this case?
- (c) Based on the provided regression euation predict the 'Total Fertility Rate' total fertility rate of a country for which the corresponding 'Adult Female Literacy' is 91.5%.

Ans:

7.

Consider a data set containing two continious variables, the Adult Female Literacy rate and the Average Life Expectancy of a sample of 35 countries. For a regression model 'Average Life Expectancy' is considered to be the response variable (Y) while the corresponding Adult Female Literacy rate (X) is used as a covariate. The following summary of the data is provided: worldwide.

$\bar{X}$	$\bar{Y}$	S_X	S_Y	S_{XY}
81.778	71.283	22.578	7.312	139.886

Finally we consider a simple linear regression model: $\hat{Y} = a + bX$ where a and b denotes the intercept and the slope corrrespondingly. Based on the information provided, answer the following questions:

(a) Compute the correlation between the variables 'Average Life Expectancy' and ' Adult Female Literacy rate'.

(b) Compute the value of the **slope** and its **interpretation**.

(c) Compute the value of the intercept and explain the corresponding interpretation. Is the interpretation meaningful in this case?

(d) Based on the provided regression euation predict the 'Average Life Expectancy' of a country for which the corresponding 'Adult Female Literacy' is 85%.

Ans:

8.

Suppose that the management of a chain of package delivery stores would like to investigate if the weekly sales (in thousands of dollars) for individual stores is based on the number of customers who made purchases. A random sample of 50 stores was selected from among all the stores in the chain and the weekly sales and the number of customers made a purchase is recorded and regression equation is calculated to be:

$$\hat{Y} = 15 + 2.5X$$

$\bar{X}$	$\bar{Y}$	S_X	S_Y	S_{XY}
81.778	71.283	22.578	7.312	139.886

Finally we consider a simple linear regression model: $\hat{Y} = a + bX$ where a and b denotes the intercept and the slope corrrespondingly. Based on the information provided, answer the following questions:

(a) Compute the correlation between the variables 'Average Life Expectancy' and ' Adult Female Literacy rate'.

(b) Compute the value of the **slope** and its **interpretation**.

(c) Compute the value of the intercept and explain the corresponding interpretation. Is the interpretation meaningful in this case?

(d) Based on the provided regression euation predict the 'Average Life Expectancy' of a country for which the corresponding 'Adult Female Literacy' is 85%.

Ans: