Assignment 02: Deadline December 30, 2024

1. Interesting data are given as:

X	76	77	78	79	80	80	81	82	83	84	85	86	88
Y	35	31	40	25	32	36	29	34	38	26	32	28	27

a. **a.** Sketch a scatter plot **b.** Compute the correlation coefficient, r. **c**. Find the regression equation
d. Compute Spearman rank correlation

2. Some data are given as:

X	4	3	5	6	3	8	9
Y	35	28	44	40	22	61	82

a. Sketch a scatter plot. **b.** Compute the correlation coefficient, r **c.** estimate values of Y when X=20
d. Compute Spearman rank correlation

3. University student who waits on tables at a local restaurant recorded the cost of meals for two people (to the nearest rand) and the tip left (to the closest rand)

Meal Cost (R) (x)	60	75	53	107	49	121	65	137
Tip (R) (y)	5	6	6	13	5	10	7	13

a) Use your calculator to find the line of best fit /the regression function.

b) Using your answer from a) above, calculate how much the waiter should expect to be given as a tip if a meal costs Tsh.200.

c) Determine the correlation coefficient.

d) What can be deduced from your answer in c)?

e) Do you think that the regression function that you found in a) above would be a reliable predictor for a meal costing Tsh.400?

f) Compute Spearman rank correlation

4. The following data for 8 learners for two tests is given.

X	95	43	56	30	48	88
Y	47	72	68	78	75	54

a) Draw a scatter plot for this data.

b) Find the equation of the line of best fit using your calculator.

c) What does the sign (+/-) of the gradient tell you?

d) What mark do you predict for Richard who got 72% for test x?

e) Calculate the correlation coefficient.

f) What can be deduced from your answer in e)?

g) What mark can you expect a learner to have achieved for test x if he achieved 85 for test y?

h) Compute Spearman rank correlation

5. How many mistakes get made during data entry? The following table gives the number of mistakes made by 13 data entry clerks who enter medical data from case report forms. These forms are submitted by doctors who participate in studies of the

performance of drugs for treating various illnesses. The column entered indicates the number of values entered and the column errors gives the number of coding errors that were detected among these

Entered	1958	7749	2829	4239	3303	5706	3770	3363	1740	3404	1640	3803	1529
Errors	28	36	42	18	54	34	40	36	23	27	26	56	20

a. Make a scatter plot of these data. Which did you choose for the response and explanatory variables? Describe any patterns
b. Find correlation for these data
c. Suppose we were to record the counts in the table in hundreds, so 1958 became 19.58. How would the correlation change? Why?
d. Write a sentence or two that interprets the value of this correlation. Use language that would be understood by someone familiar with data entry rather than correlation
e. One analyst concluded, “It is clear from this correlation that clerks who enter more values make more mistakes. Evidently they become tired as they enter more values.” Explain why this explanation is not an appropriate conclusion.
b. Compute Spearman rank correlation

6. The data below are productivity measures used to grade 12 employees who work on an assembly line. Each employee was tested once, then again a month later.

First test	64	53	18	40	24	16	67	46	64	32	71	16
Second test	76	62	39	57	41	31	75	62	64	54	65	51

c. Would you expect the scores to be associated?
d. Make a scatterplot of these data and describe the relationship, if any.
e. Regression equation to estimate the productivity
f. Compute Spearman rank correlation