



# COMSATS UNIVERSITY ISLAMABAD (CUI)

DEPARTMENT OF COMPUTER SCIENCE  
SPRING 2024  
TERMINAL EXAM

Course: DSC300-Statistical Methods in Data Science  
Maximum Marks: 50  
Class: BSDS-4

Instructor: Basharat Mahmood  
Dated: June 10, 2024.

Table 1 shows a partial dataset collected for the rain prediction.

Table 1: Sample data for rain prediction

|   | Date     | Precipitation | Temperature_max | Temperature_min | Wind Speed | Rain |
|---|----------|---------------|-----------------|-----------------|------------|------|
| 1 | 1/1/2012 | 0             | 12.8            | 5               | 4.7        | 1    |
| 2 | 1/2/2012 | 10.9          | 10.6            | 2.8             | 4.5        | 1    |
| 3 | 1/3/2012 | 0.8           | 11.7            | 7.2             | 2.3        | 1    |
| 4 | 1/4/2012 | 0             | 6.1             | -1.1            | 5.1        | 0    |
| 5 | 1/5/2012 | 0             | 6.1             | -1.7            | 1.9        | 0    |
| 6 | 1/6/2012 | 0             | 5               | -2.8            | 1.3        | 0    |
| 7 | 1/7/2012 | 0             | 7.2             | 2.8             | 2.3        | 1    |
| 8 | 1/8/2012 | 0             | 10              | 2.8             | 2          | 0    |
| 9 | 1/9/2012 | 4.3           | 9.4             | 5               | 3.4        | 1    |

LO2: Apply appropriate statistical methods to make inference about population parameters using sample statistics.

Question No. 1

15 Marks

- If it is required to make an estimation about the average "Wind Speed" on rainy days, what process will be used (Explain the phases of estimation process)? [03 Marks]
- If testing of hypothesis is used instead of estimation, then what is the difference between both methods? [02 Marks]
- Based on the sample data (Table 1) construct a 99% confidence interval for the "Wind Speed" on a rainy day? [5 Marks]  
 $\alpha = 99\%$   
 $1.4902$  to  $4.6290$
- Does it can be concluded that the average "Temperature\_min" on rainy days remains higher than the average "Temperature\_min" on sunny days? (use 0.05 level of significance)? [5 Marks]

CLO 3: Apply appropriate statistical models to analyse relationship among different variables.

Question No. 2

25 Marks

- ✓ 1. Apply appropriate method to measure the relationship between "Temperature\_min" and "Rain". Also interpret your measure. [05 Marks]  $0.8137$
- ② What <sup>proportion</sup> portion of the variation in rain can be accounted for the linear relationship with "Temperature\_min". [02 Marks]  $0.6622$
- ✓ 3. Can it be concluded at 0.05 level of significance that rain is strongly related to "Temperature\_min". [05 Marks]  $\alpha = 0.05$  accept.  $X_1$   $X_2$
- ✓ 4. If "Wind Speed" is dependent on "Temperature\_min" and "Temperature\_max", then fit a multiple linear regression model to predict the "Wind Speed". [07 Marks]
5. Show that the sum of errors (residuals) in the above model is zero. [04 Marks]  $b_1 = -0.05$   $b_2 = 0.699$
- ✓ 6. Estimate the "Wind Speed" using your fitted model when "Temperature\_min" is 2.4 and "Temperature\_max" is 14. [02 Marks]  $a = 9.5210$

CLO 4: Apply decomposition techniques to analyse time-series data.

Question No. 3

10 Marks

- ✓ 1. What is time series detrending? How can it be done for additive and multiplicative models? [03 Marks]
- ✓ 2. Consider the data given in Table 1, find the trend values of "Wind Speed" by fitting a trend equation using the least-square method. [05 Marks]
- ✓ 3. De-trend the data (assuming additive model) [02 Marks]  $7$

$$Y_t - T_t = S_t + \epsilon_t$$



The End