

# UNIVERSITY OF WESTMINSTER

School of Computer Science and Engineering

**Module Code:** 6DATA005W  
**Module Title:** Advanced Analytics  
**Module Leader:** Dr Fang He  
**Exam Time:** 11:10-12:40 17<sup>th</sup> December 2021

## Instructions to Candidates:

**Please read the instructions below before starting the paper**

- The Module Leader will be available during the exam time to respond to any queries
- This is an individual piece of work so do not collude with others on your answers as this is an academic offence
- Where the University believes that academic misconduct has taken place the University will investigate the case and apply academic penalties as published in [Section 10 Academic Misconduct regulations](#).

## Module Specific Information

Candidates should attempt ALL questions. Full mark can only be obtained by correct answers to ALL questions.

### Question 1 (30 marks)

Consider the following linear program:

$$\max x_1 + 2x_2$$

$$s. t. \quad x_1 - x_2 \geq -2 \quad \checkmark$$

$$x_1 + x_2 \leq 4 \quad \checkmark$$

$$x_1 \leq 2.5 \quad \times$$

$$x_2 \leq 3 \quad \times$$

$$x_1, x_2 \geq 0$$

- a) **(10 marks)** Sketch the feasible region of the LP. Is the feasible region unbounded?
- b) **(5 marks)** Are any of the above constraints redundant? If so, indicate which one(s) and why.
- c) **(10 marks)** Solve the LP using the graphical method. Explain your approach.
- d) **(5 marks)** Is there more than one optimal solution? If so give two different solutions. If not, explain using the graphical method why not?

**Question (2) (30 marks)**

On a particular day during the tourist season a rent-a-car company must supply cars to four destinations according to the following schedule:

<i>Destination</i>	<i>Cars required</i>
A	2
B	3
C	5
D	7

The company has three branches from which the cars may be supplied. On the day in question, the inventory status of each of the branches was as follows:

<i>Branch</i>	<i>Cars available</i>
1	6
2	1
3	10

The distances between branches and destinations are given by the following table:

<i>Branch</i>	<i>Destination</i>			
	A	B	C	D
1	7	11	3	2
2	1	6	0	1
3	9	15	8	5

Plan the day's activity such that supply requirements are met at a minimum cost (assumed proportional to car-miles travelled). Please present your model for the problem by giving clear definition of decision variables **(5 marks)**, objective function **(10 marks)** and constraints **(15 marks)**.

### Question (3) (20 marks)

The data below gives the total sales of x: Pay as You Go mobile handsets and y: Pay monthly contracts obtained for 10 sales staff in a London mobile phone store for the month of October 2021

Staff Name	Sam	Jeff	Jennifer	Reshma	Danial	Tim	Lee	Julia	Oscar	Amanda
x: PAYG	20	23	8	29	14	11	11	20	17	17
y: Contracts	30	35	21	33	33	26	22	31	33	36

- By applying a scatter plot diagram and the mean formula to the above table, is there a connection (association) between the sales achieved by the ten members of staff for PAYG and Contracts? Justify (describe) your answer in the context of the variables mean and the plotted scatter graph. (8 Marks)
- Calculate the covariance  $\text{cov}(x,y)$  then describe the tendency of the trend based on the obtained calculation. (4 marks)
- Calculate the Pearson's product moment correlation coefficient  $r$  and describe the association between both PAYG and Contracts sales in relation to  $r$ . (4 marks)
- List two limitation of Pearson's product moment correlation. (4 marks)

#### Question (4) (20 marks)

Below is an extraction from a KPI (Key Performance Indicator) data sheet for a mobile phone retailer store in London in 2021. The number of converted customers leads per month  $x$  over the period March to October and the monthly gross profit  $y$  in £1000 for each month.

Month	March	April	May	June	July	Aug	Sept	Oct
x: Conversions per month	50	100	150	200	250	300	350	400
y: Gross Profit in £1000	37	48	60	71	80	90	102	109

- Using the method of least squares, calculate the gradient and the intercept of the linear regression line. Write the linear regression 'y' on 'x' line equation. (6 Marks)
- With the linear regression method, predict the store's gross profit using the 'y' on 'x' line for the months November and December 2021 for targeted conversions rates of 375 and 1000 respectively. (6 Marks)
- Taking into account the linear regression line from (b), What is the difference between interpolating and extrapolating the store's gross profit based on leads conversions. For each scenario give an example value. (8 Marks)

## Formula sheet

### Covariance

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

which is known as the **covariance** and denoted by  $\text{cov}(X,Y)$  or  $s_{xy}$ . For shorthand it is normally written as

$$\frac{1}{n} \sum (x - \bar{x})(y - \bar{y})$$

Thus

$$\frac{1}{n} \sum (x - \bar{x})(y - \bar{y}) = \frac{1}{n} \sum xy - \bar{x}\bar{y}$$

### Pearson's product moment correlation coefficient

$$r = \frac{\frac{1}{n} \sum xy - \bar{x}\bar{y}}{s_x s_y}$$

where  $s_x = \sqrt{\frac{1}{n} \sum x^2 - \bar{x}^2}$  and  $s_y = \sqrt{\frac{1}{n} \sum y^2 - \bar{y}^2}$ .

### Linear Regression

$$\bar{y} = \frac{\sum y}{n}, \quad \bar{x} = \frac{\sum x}{n}$$

$$y - \bar{y} = \frac{s_{xy}}{s_x^2} (x - \bar{x})$$

**END OF EXAM PAPER**