Statistics 17 – PMCC

Please **complete** this homework by _____. Start it early. If you can't do a question you will then have time to ask your teacher for help or go to a drop in session.

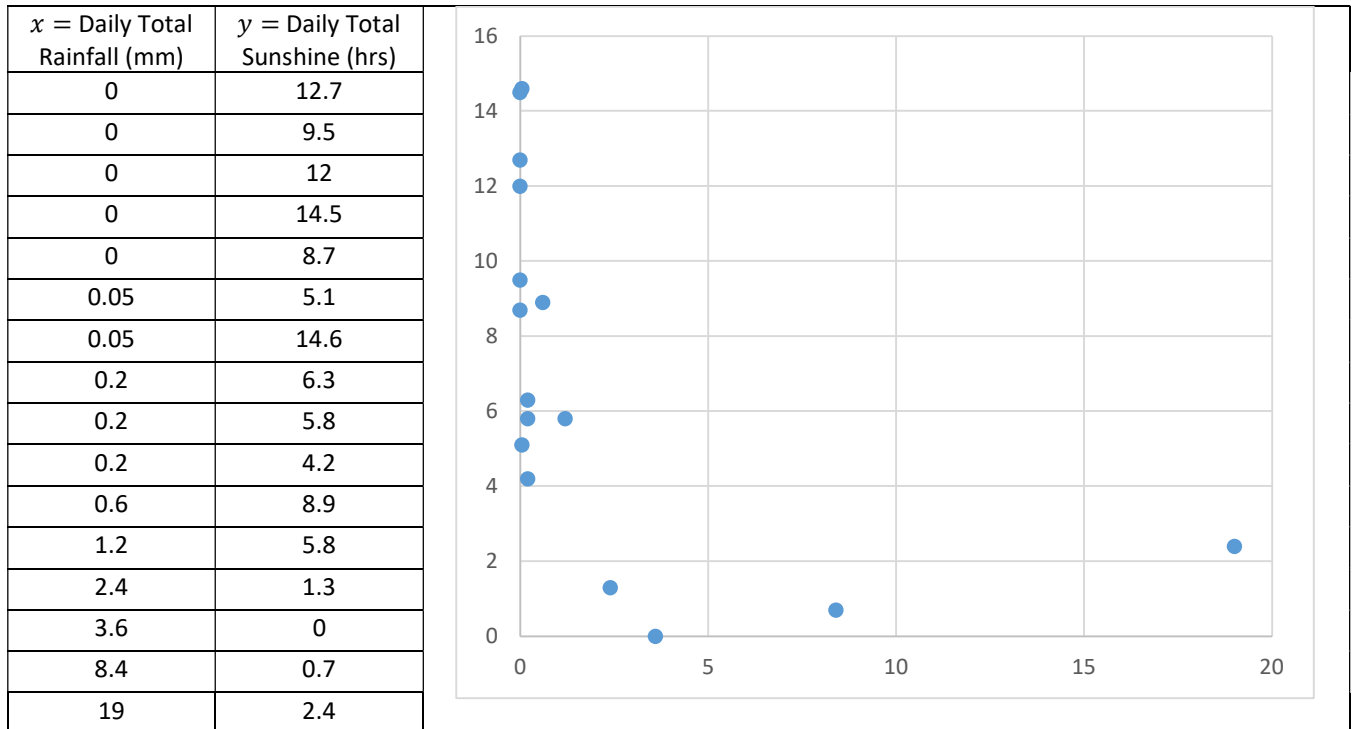## Section 1 – Review of previous topics.
Please complete all questions.

1. From studying the large data set for Hurn in 1987 Michelle calculates that the probability of a "light" wind speed (as measured by the Beaufort conversion) is 0.9

a) Asma takes a random sample of 20 days from the large data set for Hurn in 1987. Using Michelle's probability calculate:
    i.   The probability that less than 16 days have a "light" wind speed
    ii. The probability that more than 15 days have a "light" wind speed

b) Asma now thinks that Michelle's calculations are wrong and so carries out a 5% hypothesis test.
    i.   Determine the critical region
    ii. Given that 16 days from Asma's sample have a "light" wind speed, what conclusion should she make?

2. There are three sets of traffic lights on Abdul's journey to work. The independent probabilities that Abdul has to stop at the first, second and third set of lights are 0.2, 0.3 and 0.1 respectively.

a) Find the probability that Abdul has to stop at each of the first two sets of lights but does not have to stop at the third set

b) Find the probability that Abdul has to stop at exactly two of the three sets of lights

c) Find the probability that Abdul has to stop at the first set of lights, given that he has to stop at exactly two sets of lights

3. Two events A and B are such that $P(A) = 0.3, P(A \cap B) = 0.1, P(A/B') = 0.4$. Find:

a) $P(A'/B')$

b) $P(B/A)$

4. Charles takes a random sample of 16 days, from the large data set, to find a relationship between the Daily Rainfall and Daily Sunshine in Cambourne in 2015.

x = Daily Total Rainfall (mm)	y = Daily Total Sunshine (hrs)
0	12.7
0	9.5
0	12
0	14.5
0	8.7
0.05	5.1
0.05	14.6
0.2	6.3
0.2	5.8
0.2	4.2
0.6	8.9
1.2	5.8
2.4	1.3
3.6	0
8.4	0.7
19	2.4

a) Calculate the value of r
b) By looking at the data how do you think Charles interpreted a "trace" of rain
c) Charles found the regression equation to be $y = 8.13 - 0.49x$. Find the range of values of x for which this equation gives sensible values of y.
d) Explain why an equation of the form $y = ax^b$ is more likely to be appropriate
e) Explain what we can deduce about the value of b from the graph above
f) Why is it not possible to use coded values to determine the values of a or b

5. Without using any data, or putting scales on your axes, sketch a frequency polygon that you would expect for a very large sample of the following data:
a) Height of people in the UK today
b) Average daily windspeed in Heathrow
c) IQ of people in a country

6. How could a frequency polygon be used to find probabilities, if the total area under the graph is equal to 1

Section 2 – Consolidation of this week's topic.
Please <u>complete</u> all questions.

1.

a) State the hypotheses to determine whether there is evidence of positive correlation

(2 marks)

b) Copy and complete the table below to determine whether there is evidence of positive correlation in each scenario

	Sample size	Correlation coefficient	Significance level	Critical value	Conclusion: Accept H_0 or Accept H_1
i.	10	0.5	5%		
ii.	20	0.5	5%		
iii.	30	0.5	10%		
iv.	40	0.4	10%		

(8 marks)

2. Copy and complete the table below to carry out hypothesis tests for correlation

	H_1	Sample size	Correlation coefficient	Significance level	Critical region(s)	Conclusion: Accept H_0 or Accept H_1
a)	$\rho < 0$	20	-0.4	5%		
b)	$\rho \neq 0$	30	0.5	2%		
c)	$\rho \neq 0$	40	-0.3	1%		

(6 marks)

3. Charles uses the large data set for the first 10 days of May 1987 to assess whether there is linear correlation between the Daily Mean Pressure and Daily Mean Wind Speed in Beijing

Daily Mean Pressure (hPa)	Daily Mean Wind Speed
1016	4.0
1018	4.5
1018	7.5
1016	8.0
1012	3.5
1007	9.5
1010	5.5
1012	8.5
1015	4.5
1015	6.5

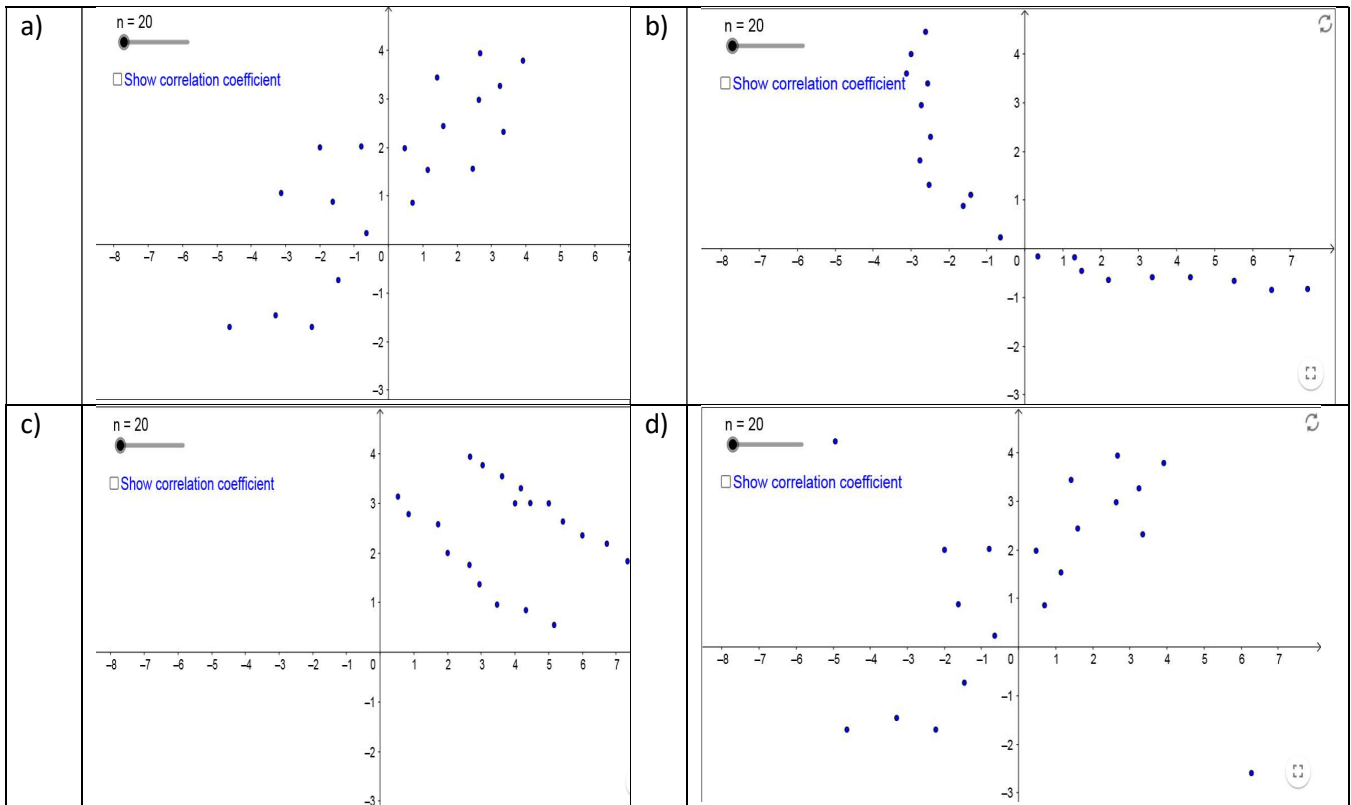a) Calculate the value of r (2 marks)

b) Test at a 0.1 significance level for:

   i.   linear correlation   (3 marks)

   ii.   positive linear correlation   (3 marks)

  iii.   negative linear correlation   (3 marks)

c) What unit is mean wind speed measured in   (1 mark)

d) Is it possible that the answers to b) would be different if another unit for wind speed was used? Explain your answer   (2 marks)

4. Asma calculates the value of the correlation coefficient to be -0.4367. She carries out a 5% hypothesis test and concludes "Accept H_1".

What is the smallest possible sample size Asma used if:
a) she was testing for negative linear correlation   (3 marks)
b) she was testing for any linear correlation   (3 marks)

5. Estimate, to one decimal place, the value of the correlation coefficient for the following scatter diagrams.   (4 marks)

a)	b)
c)	d)

Total: 40 Marks