

Statistics 16 – Linearising

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.

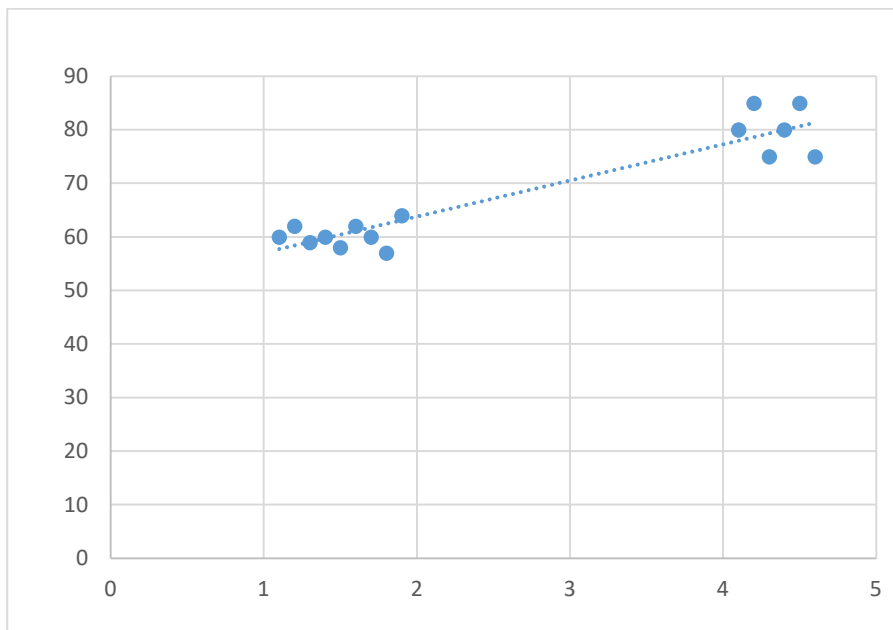
Please complete all questions.

1. A random sample of data from the 2015 sections of the large data set is used to find the relationship between C , the daily mean total cloud cover (oktas) and V , the daily mean visibility (km). The equation of the regression line is found to be $V = 69.62 - 7.31C$
 - a) Interpret, in context, the constants in the equation above
 - b) What visibility would you expect on:
 - i. The 31st May when the cloud cover is recorded as being 4 oktas
 - ii. The 1st December when the cloud cover is recorded as being 6 oktas
 - c) Which answer from b) is the most dubious? Give reasons
 - d)
 - i. Find the value of V , when $C = 9$
 - ii. Explain why the answer to d) does not tell us anything about the visibility

2. Two events A and B are such that $P(A)=0.6$, $P(B)=0.5$ and $P(A \cup B)=0.85$.
 - a) Represent this information in a 2 way table
 - b) Find $P(B/A)$
 - c) Find $P(A'/B')$

3. Each of the 30 students in a class play at least one of squash, hockey and tennis.
 - 18 students play squash
 - 19 students play hockey
 - 17 students play tennis
 - 8 students play squash and hockey
 - 9 students play hockey and tennis
 - 11 students play squash and tennis
 - a) Represent this information in a Venn diagram
 - b) A student playing squash is picked from the class. Find the probability that this student does not play hockey
 - c) Two different students are picked from the class. Given that the first student plays squash, find the probability that the second student does not play squash.

4. A dog food company claim that 8 out of 10 owners prefer their product to other brands. 40 owners were surveyed and a 5% hypothesis test carried out to determine whether there was any significant evidence that the company was overestimating the popularity of its product.
- Determine the critical region
 - State the conclusion if 31 owners said that their dogs preferred the product.
5. A geyser is a hot spring which erupts from time to time. The duration of each eruption, x minutes, and the waiting time until the next eruption, y minutes, are recorded. For a random sample of 15 eruptions it is calculated that there is very strong positive linear correlation between x and y
- Michelle claims that this value of r shows that a longer eruption causes a delay for the next eruption. Comment on Michelle's claim.



- Make two further comments given the data and scatter diagram above

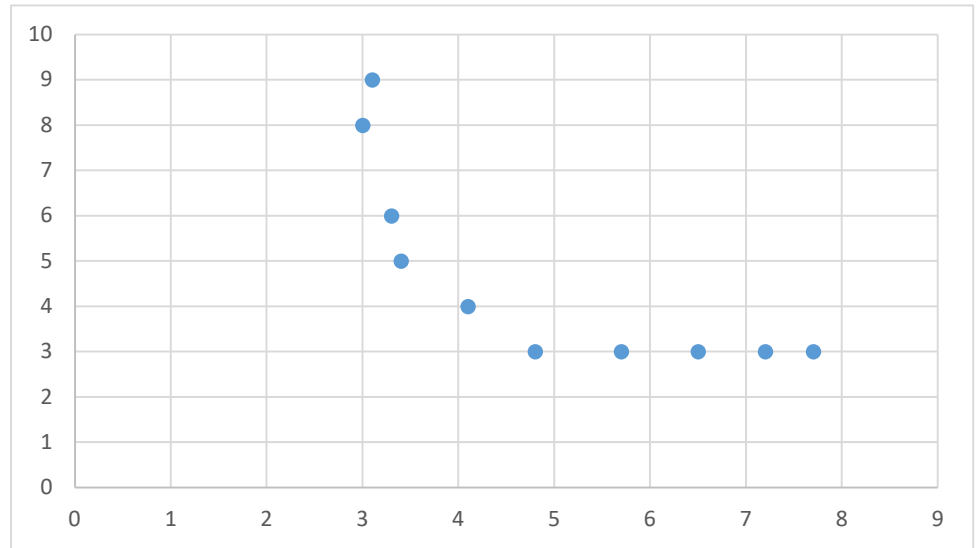
Section 2 – Consolidation of this week's topic.

Please complete all questions.

1. The relationship between P , the population of bacteria and t , the time (hours) after 12:00 is thought to be of the form $P = ab^t$ where a and b are constants. A graph of $y = \log_{10} P$ against $x = t$ is plotted and shows very strong positive linear correlation
- Given that the regression line on the graph goes through the points (0, 2) and (5, 5), find the values of a and b (4 marks)
 - Hence find
 - the amount of bacteria at 12:00 (1 mark)
 - the period of time, in minutes in which the bacteria will have increased to three times its initial size (3 marks)

2. Charles uses the large data set to investigate the relationship between S , the daily total sunshine and W , the daily mean wind speed in Leeming in 1987. He takes a random sample of 10 days, which is shown below, alongside a scatter diagram.

Daily Total Sunshine (hrs)	Daily Mean Windspeed (kn)
4.8	3
5.7	3
6.5	3
7.2	3
7.7	3
4.1	4
3.4	5
3.3	6
3	8
3.1	9



- Which variable could be described as the independent variable? (1 mark)
- Why would it be unwise to find the equation of the regression line? (1 mark)
- Abdul thinks the relationship may be of the form $W = aS^b$ where a and b are constants. He plots a new scatter diagram changing the axes appropriate using base 10. What does he plot on each axis (3 marks)
- The equation of the regression line for Abdul's idea is $y = 1.35 - 1.1x$. Hence find the equation that Abdul wants to use for W in terms of S (5 marks)
- Approximate the daily mean windspeed when the daily total sunshine is 6 hours (1 mark)
- Charles doubts the relationship Abdul has found. What should Charles do to understand the relationship between S , and W , in Leeming in 1987, more confidently. (1 mark)

Total: 20 Marks