1. A meteorologist believes that there is a relationship between the daily mean windspeed, w kn, and the daily mean temperature, t °C. A random sample of 9 consecutive days is taken from past records from a town in the UK in July and the relevant data is given in the table below.

t	13.3	16.2	15.7	16.6	16.3	16.4	19.3	17.1	13.2
w	7	11	8	11	13	8	15	10	11

The meteorologist calculated the product moment correlation coefficient for the 9 days and obtained $r = 0.609$

(a) Explain why a linear regression model based on these data is unreliable on a day when the mean temperature is 24 °C
(1)

(b) State what is measured by the product moment correlation coefficient.
(1)

(c) Stating your hypotheses clearly test, at the 5% significance level, whether or not the product moment correlation coefficient for the population is greater than zero.
(3)

Using the same 9 days a location from the large data set gave $\bar{t} = 27.2$ and $\bar{w} = 3.5$

(d) Using your knowledge of the large data set, suggest, giving your reason, the location that gave rise to these statistics.
(1)

a) Linear Regression requires us to extrapolate data.
We know that we have to extrapolate since no data is given for 24°C. We know that extrapolating can be unreliable, hence a linear regression will be unreliable (1)

b) Product Moment Correlation coefficient measures the linear association between two variables.
In our case, we have the linear association between daily mean windspeed, w and the daily mean temperature, t. (1)

c) $\alpha = 0.05$

One sided Test

* Critical Value *

• Product moment correlation data table in the formula sheet.

$H_0 : \rho = 0$ v.s. $H_1 : \rho > 0$ 1

$n = 9$

$\alpha = 0.05$ => From table, the critical value = 0.5822 1

At the begining, we were told that $r = 0.609$.

=> $r = 0.609 > 0.5822$ => We reject H_0, and this means that we can conclude that the product moment correlation coefficient is greater than zero. 1

d)

Large data set:

• this will not be the UK since the temperature is too high
• Options are Perth, Beijing or Jacksonville
• exclude Perth since its July, it will be winter there and thus it won't be that hot
• Wind is low (3.5) => Not near the sea.

=> We suggest that the location is Beijing. 1

2. Tessa owns a small clothes shop in a seaside town. She records the weekly sales figures, £w, and the average weekly temperature, t °C, for 8 weeks during the summer.
The product moment correlation coefficient for these data is −0.915

$\rightarrow 0.05$

(a) Stating your hypotheses clearly and using a 5% level of significance, test whether or not the correlation between sales figures and average weekly temperature is negative.
(3)

(b) Suggest a possible reason for this correlation.
(1)

Tessa suggests that a linear regression model could be used to model these data.

(c) State, giving a reason, whether or not the correlation coefficient is consistent with Tessa's suggestion.
(1)

(d) State, giving a reason, which variable would be the explanatory variable.
(1)

Tessa calculated the linear regression equation as $w = 10\,755 - 171t$

(e) Give an interpretation of the gradient of this regression equation.
(1)

a) $H_0: \rho = 0 \qquad H_1: \rho < 0$ - (1)

$n = 8$, significance level $= 5\%$ → Use these to find critical value from the table of critical values of the product moment correlation coefficient

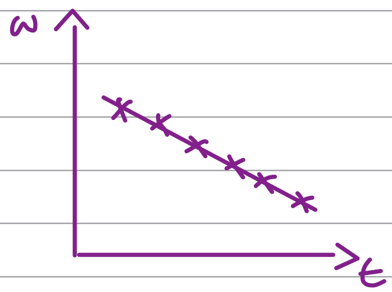
critical value: -0.6215 - (1)

critical reigon: $r < -0.6215$

$-0.915 < -0.6215$

We can reject Ho - there is evidence of a negative correlation between w and t. - (1)

b) When temperature increases, people want to spend more time outside (e.g. on the beach) instead of shopping. - (1)

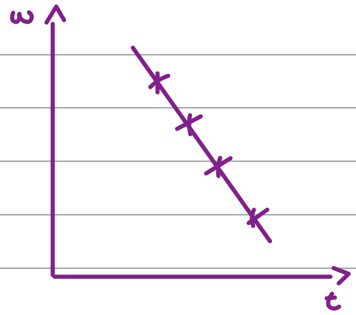
c) 

$\Rightarrow r = 1$

The correlation coefficient is consistent - as -0.915 is very close to -1. - (1)

d) Temperature would be the explanatory value, as sales are likely to depend on temperature. - (1)

e)



Every degree rise in temperature leads to a drop in weekly sales of £171. – 1

3. Barbara is investigating the relationship between average income (GDP per capita), x US dollars, and average annual carbon dioxide (CO_2) emissions, y tonnes, for different countries.

She takes a random sample of 24 countries and finds the product moment correlation coefficient between average annual CO_2 emissions and average income to be 0.446

(a) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not the product moment correlation coefficient for all countries is greater than zero.

(3)

Barbara believes that a non-linear model would be a better fit to the data.
She codes the data using the coding $m = \log_{10}x$ and $c = \log_{10}y$ and obtains the model $c = -1.82 + 0.89m$

The product moment correlation coefficient between c and m is found to be 0.882

(b) Explain how this value supports Barbara's belief.

(1)

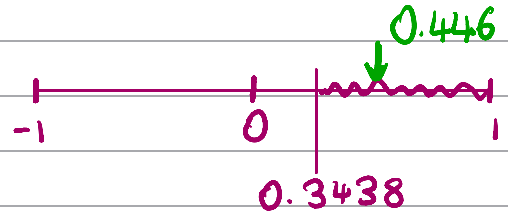
(c) Show that the relationship between y and x can be written in the form $y = ax^n$ where a and n are constants to be found.

(5)

a) $H_0: \rho = 0$ $n = 24$

$H_1: \rho > 0$ (1) $SL = 5\%$

Critical value is 0.3438 (1)



$0.446 > 0.3438$ so there is evidence that the pmcc is greater than 0 (1)

Product Moment Coefficient					
Level					Sample size, n
0.10	0.05	0.025	0.01	0.005	
0.8000	0.9000	0.9500	0.9800	0.9900	4
0.6870	0.8054	0.8783	0.9343	0.9587	5
0.6084	0.7293	0.8114	0.8822	0.9172	6
0.5509	0.6694	0.7545	0.8329	0.8745	7
0.5067	0.6215	0.7067	0.7887	0.8343	8
0.4716	0.5822	0.6664	0.7498	0.7977	9
0.4428	0.5494	0.6319	0.7155	0.7646	10
0.4187	0.5214	0.6021	0.6851	0.7348	11
0.3981	0.4973	0.5760	0.6581	0.7079	12
0.3802	0.4762	0.5529	0.6339	0.6835	13
0.3646	0.4575	0.5324	0.6120	0.6614	14
0.3507	0.4409	0.5140	0.5923	0.6411	15
0.3383	0.4259	0.4973	0.5742	0.6226	16
0.3271	0.4124	0.4821	0.5577	0.6055	17
0.3170	0.4000	0.4683	0.5425	0.5897	18
0.3077	0.3887	0.4555	0.5285	0.5751	19
0.2992	0.3783	0.4438	0.5155	0.5614	20
0.2914	0.3687	0.4329	0.5034	0.5487	21
0.2841	0.3598	0.4227	0.4921	0.5368	22
0.2774	0.3515	0.4133	0.4815	0.5256	23
0.2711	0.3438	0.4044	0.4716	0.5151	24
0.2653	0.3365	0.3961	0.4622	0.5052	25
0.2598	0.3297	0.3882	0.4534	0.4958	26
0.2546	0.3233	0.3809	0.4451	0.4869	27

use tables to find critical value for part a

b) The value is closer to 1 (1)

c) $c = -1.82 + 0.89m$

$$\log_{10} y = -1.82 + 0.89 \log_{10} x \quad 1$$

$$\log_{10} y = 10^{-1.82 + 0.89(\log_{10} x)} \qquad 10^{\log_{10} A} = A$$

$$y = 10^{-1.82 + 0.89(\log_{10} x)}$$

1

$$x^a \times x^b = x^{a+b}$$

$$y = 10^{-1.82} \times 10^{0.89(\log_{10} x)} \qquad \log_{10} x^a = a \log_{10} x$$

1

$$y = 0.015 \times 10^{(\log_{10} x)^{0.89}} \qquad 10^{\log_{10} A} = A$$

$$y = 0.015 x^{0.89}$$

2

4. A random sample of 15 days is taken from the large data set for Perth in June and July 1987.

The scatter diagram in Figure 1 displays the values of two of the variables for these 15 days.

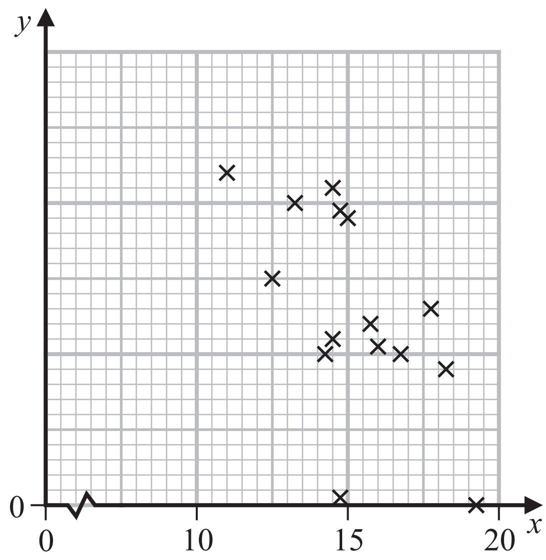


Figure 1

(a) Describe the correlation. Negative ①

(1)

The variable on the x-axis is Daily Mean Temperature measured in °C. ① rainfall or pressure

(b) Using your knowledge of the large data set, Rainfall → mm ① corresponding unit

(i) suggest which variable is on the y-axis, → Pressure → hPa, Pascals, hectopascals, mb, or Millibars

(ii) state the units that are used in the large data set for this variable.

(2)

Stav believes that there is a correlation between Daily Total Sunshine and Daily Maximum Relative Humidity at Heathrow.

He calculates the product moment correlation coefficient between these two variables for a random sample of 30 days and obtains $r = -0.377$

(c) Carry out a suitable test to investigate Stav's belief at a 5% level of significance. State clearly

- your hypotheses
- your critical value

(3)

On a random day at Heathrow the Daily Maximum Relative Humidity was 97%

(d) Comment on the number of hours of sunshine you would expect on that day, giving a reason for your answer.

(1)

c) let ρ be the population correlation coefficient

$H_0: \rho = 0$ 1

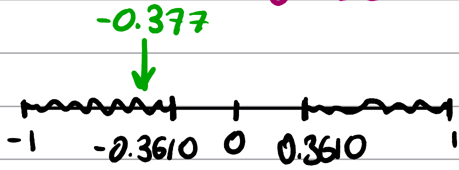
$H_1: \rho \neq 0$

$n = 30$

$SL = 5\%$

two-tailed test so $\div 2$

$\therefore 0.025$



Critical value $= -0.3610$ 1

$r = -0.377 < -0.3610$ 1

$\therefore$ Significant result and there is evidence of a correlation between Daily total Sunshine and Daily Max. Relative humidity

Product Moment Coefficient					Sample size, n
Level					
0.10	0.05	0.025	0.01	0.005	
0.8000	0.9000	0.9500	0.9800	0.9900	4
0.6870	0.8054	0.8783	0.9343	0.9587	5
0.6084	0.7293	0.8114	0.8822	0.9172	6
0.5509	0.6694	0.7545	0.8329	0.8745	7
0.5067	0.6215	0.7067	0.7887	0.8343	8
0.4716	0.5822	0.6664	0.7498	0.7977	9
0.4428	0.5494	0.6319	0.7155	0.7646	10
0.4187	0.5214	0.6021	0.6851	0.7348	11
0.3981	0.4973	0.5760	0.6581	0.7079	12
0.3802	0.4762	0.5529	0.6339	0.6835	13
0.3646	0.4575	0.5324	0.6120	0.6614	14
0.3507	0.4409	0.5140	0.5923	0.6411	15
0.3383	0.4259	0.4973	0.5742	0.6226	16
0.3271	0.4124	0.4821	0.5577	0.6055	17
0.3170	0.4000	0.4683	0.5425	0.5897	18
0.3077	0.3887	0.4555	0.5285	0.5751	19
0.2992	0.3783	0.4438	0.5155	0.5614	20
0.2914	0.3687	0.4329	0.5034	0.5487	21
0.2841	0.3598	0.4227	0.4921	0.5368	22
0.2774	0.3515	0.4133	0.4815	0.5256	23
0.2711	0.3438	0.4044	0.4716	0.5151	24
0.2653	0.3365	0.3961	0.4622	0.5052	25
0.2598	0.3297	0.3882	0.4534	0.4958	26
0.2546	0.3233	0.3809	0.4451	0.4869	27
0.2497	0.3172	0.3739	0.4372	0.4785	28
0.2451	0.3115	0.3673	0.4297	0.4705	29
0.2407	0.3061	0.3610	0.4226	0.4629	30
0.2070	0.2638	0.3120	0.3665	0.4026	40
0.1843	0.2353	0.2787	0.3281	0.3610	50

d)

Humidity is high and there is evidence of correlation 1 and $r < 0$ so would expect lower than average amount of sunshine

5. Marc took a random sample of 16 students from a school and for each student recorded

- the number of letters, x, in their last name
- the number of letters, y, in their first name

His results are shown in the scatter diagram on the next page.

(a) Describe the correlation between x and y.

(1)

Marc suggests that parents with long last names tend to give their children shorter first names.

(b) Using the scatter diagram comment on Marc's suggestion, giving a reason for your answer.

(1)

The results from Marc's random sample of 16 observations are given in the table below.

x	3	6	8	7	5	3	11	3	4	5	4	9	7	10	6	6
y	7	7	4	4	6	8	5	5	8	4	7	4	5	5	6	3

(c) Use your calculator to find the product moment correlation coefficient between x and y for these data.

(1)

(d) Test whether or not there is evidence of a negative correlation between the number of letters in the last name and the number of letters in the first name.

You should

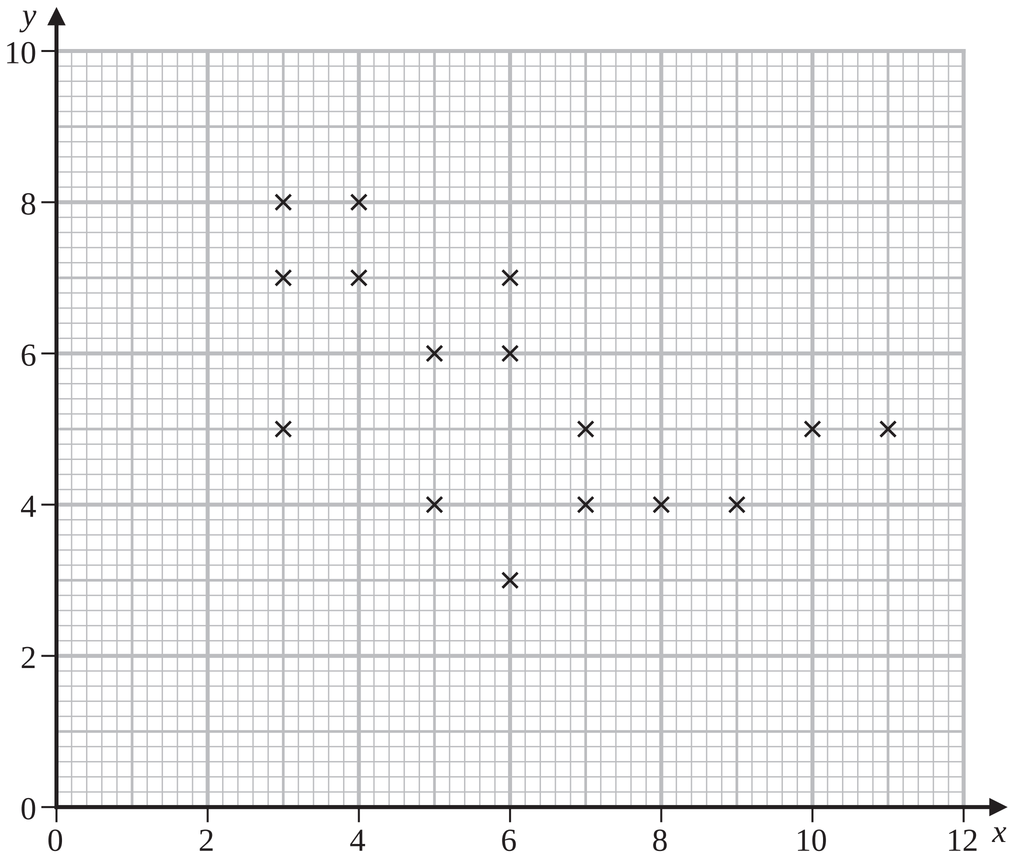
- state your hypotheses clearly
- use a 5% level of significance

(3)

a) (Slight/weak) Negative correlation. 1

b) Marc's suggestion is compatible with the data because as x (last name length) increases, y (first name length) decreases 1

c) $r = -0.545$ (3.s.f) 1



d) $H_0 : \rho = 0$ $H_1 : \rho < 0$ (1) ← $\rho < 0$ is a negative correlation, which is what we are testing for

At 5% significance, the one-tail critical value is ± 0.4259. (1)

Since our calculated value $r = -0.545$ is more extreme than ± 0.4259, this is a significant result and Ho is rejected.

Must then conclude :

∴ There is evidence to suggest a negative correlation between the number of letters in a student's last name and their first name. (1)