

A LEVEL MATHEMATICS – STATISTICS

REGRESSION AND CORRELATION (PREP)

NAME Worked Solutions

Critical Values for Correlation Coefficients

These tables concern tests of the hypothesis that a population correlation coefficient ρ is 0. The values in the tables are the minimum values which need to be reached by a sample correlation coefficient in order to be significant at the level shown, on a one-tailed test.

Product Moment Coefficient					Sample size, n	Spearman's Coefficient		
0.10	0.05	Level 0.025	0.01	0.005		Level 0.05	0.025	0.01
0.8000	0.9000	0.9500	0.9800	0.9900	4	1.0000	—	—
0.6870	0.8054	0.8783	0.9343	0.9587	5	0.9000	1.0000	1.0000
0.6084	0.7293	0.8114	0.8822	0.9172	6	0.8286	0.8857	0.9429
0.5509	0.6694	0.7545	0.8329	0.8745	7	0.7143	0.7857	0.8929
0.5067	0.6215	0.7067	0.7887	0.8343	8	0.6429	0.7381	0.8333
0.4716	0.5822	0.6664	0.7498	0.7977	9	0.6000	0.7000	0.7833
0.4428	0.5494	0.6319	0.7155	0.7646	10	0.5636	0.6485	0.7455
0.4187	0.5214	0.6021	0.6851	0.7348	11	0.5364	0.6182	0.7091
0.3981	0.4973	0.5760	0.6581	0.7079	12	0.5035	0.5874	0.6783
0.3802	0.4762	0.5529	0.6339	0.6835	13	0.4835	0.5604	0.6484
0.3646	0.4575	0.5324	0.6120	0.6614	14	0.4637	0.5385	0.6264
0.3507	0.4409	0.5140	0.5923	0.6411	15	0.4464	0.5214	0.6036
0.3383	0.4259	0.4973	0.5742	0.6226	16	0.4294	0.5029	0.5824
0.3271	0.4124	0.4821	0.5577	0.6055	17	0.4142	0.4877	0.5662
0.3170	0.4000	0.4683	0.5425	0.5897	18	0.4014	0.4716	0.5501
0.3077	0.3887	0.4555	0.5285	0.5751	19	0.3912	0.4596	0.5351
0.2992	0.3783	0.4438	0.5155	0.5614	20	0.3805	0.4466	0.5218
0.2914	0.3687	0.4329	0.5034	0.5487	21	0.3701	0.4364	0.5091
0.2841	0.3598	0.4227	0.4921	0.5368	22	0.3608	0.4252	0.4975
0.2774	0.3515	0.4133	0.4815	0.5256	23	0.3528	0.4160	0.4862
0.2711	0.3438	0.4044	0.4716	0.5151	24	0.3443	0.4070	0.4757
0.2653	0.3365	0.3961	0.4622	0.5052	25	0.3369	0.3977	0.4662
0.2598	0.3297	0.3882	0.4534	0.4958	26	0.3306	0.3901	0.4571
0.2546	0.3233	0.3809	0.4451	0.4869	27	0.3242	0.3828	0.4487
0.2497	0.3172	0.3739	0.4372	0.4785	28	0.3180	0.3755	0.4401
0.2451	0.3115	0.3673	0.4297	0.4705	29	0.3118	0.3685	0.4325
0.2407	0.3061	0.3610	0.4226	0.4629	30	0.3063	0.3624	0.4251
0.2070	0.2638	0.3120	0.3665	0.4026	40	0.2640	0.3128	0.3681
0.1843	0.2353	0.2787	0.3281	0.3610	50	0.2353	0.2791	0.3293
0.1678	0.2144	0.2542	0.2997	0.3301	60	0.2144	0.2545	0.3005
0.1550	0.1982	0.2352	0.2776	0.3060	70	0.1982	0.2354	0.2782
0.1448	0.1852	0.2199	0.2597	0.2864	80	0.1852	0.2201	0.2602
0.1364	0.1745	0.2072	0.2449	0.2702	90	0.1745	0.2074	0.2453
0.1292	0.1654	0.1966	0.2324	0.2565	100	0.1654	0.1967	0.2327

Question 1

Amar is studying the flight of a bird from its nest.

He measures the bird's height above the ground, h metres, at time t seconds for 10 values of t

Amar finds the equation of the regression line for the data to be $h = 38.6 - 1.28t$

(a) Interpret the gradient of this line.

For every second, height of bird above ground decreases by 1.28m. (1)

The product moment correlation coefficient between h and t is -0.510

(b) Test whether or not there is evidence of a negative correlation between the height above the ground and the time during the flight.

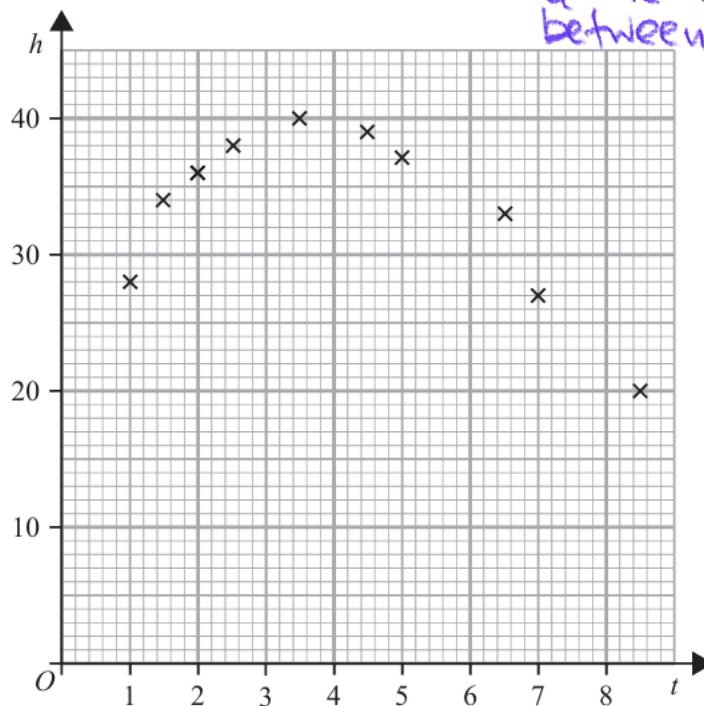
You should

- state your hypotheses clearly
- use a 5% level of significance
- state the critical value used

$H_0: \rho = 0, H_1: \rho < 0$
at 5%: $CV = -0.5494$

Since $-0.510 > -0.5494$ insufficient evidence to suggest a -ve correlation between height and time. (3) H_0 accept

Jane draws the following scatter diagram for Amar's data.



(c) With reference to the scatter diagram, state, giving a reason, whether or not the regression line $h = 38.6 - 1.28t$ is an appropriate model for these data.

The points suggest a non-linear relationship so not an appropriate model. (1)

Jane suggests an improved model using the variable $u = (t - k)^2$ where k is a constant.

She obtains the equation $h = 38.1 - 0.78u$

* (d) Choose a suitable value for k to write Jane's improved model for h in terms of t only.

$$u = (t - k)^2 \Rightarrow h = 38.1 - 0.78(t - 3.5)^2 \quad (1)$$

Question 2

Kay is studying the variables Daily Total Sunshine (x) and Daily Total Rainfall (y) from the large data set for Leeming in 2015

Kay starts with 5th May and then selects every 10th day thereafter.

- (a) State the name of the sampling technique Kay uses.

Systematic sampling

(1)

Kay wants to find the regression line of y on x for these data.

- (b) Using your knowledge of the large data set, explain how Kay might need to clean these data before finding the equation of the regression line.

Replace the 'tr' values contained with suitable value i.e 0.

The equation of the regression line Kay finds is $y = 0.741 + 0.199x$

- (c) Using your knowledge of the large data set,

(i) state the units of the gradient of the regression line, mm/hr

(ii) give an interpretation of the y -intercept of the regression line.

mean amount of rain when there is no sunshine

(2)

Kay's teacher claimed that the greater the amount of sunshine in a day the lower the amount of rain there should be.

- (d) State, giving a reason, whether the teacher's claim is true for Kay's data.

Regression line suggest positive correlation, yet the teacher claim suggest a negative correlation

(1)

The teacher used all the data for these variables from the large data set for Leeming in 2015, as a sample.

The teacher calculated the product moment correlation coefficient for x and y to be -0.160

In a suitable test to determine whether there is evidence to support the teacher's claim, the p -value was 0.015

- (e) Using a 5% level of significance, state the hypotheses and conclusion for this test.

(2)

- (f) For the test in part (e) describe

(i) the sample, Rainfall and Sunshine for Leeming in May to Oct 2015

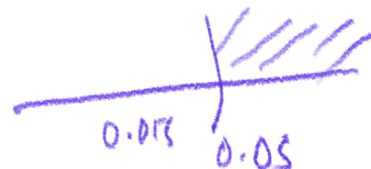
(ii) a possible population. Rainfall and Sunshine for all locations across all time periods.

(2)

(e) $H_0: \rho = 0$ $H_1: \rho < 0$

$p = 0.015 < 0.05$

Reject H_0



The test support teacher claim of a negative correlation.

Question 3

Nat is studying the displacement, y cm, of an object during a period of 7.5 seconds.

The scatter diagram in Figure 1 shows the results for 11 measurements of y at x seconds.

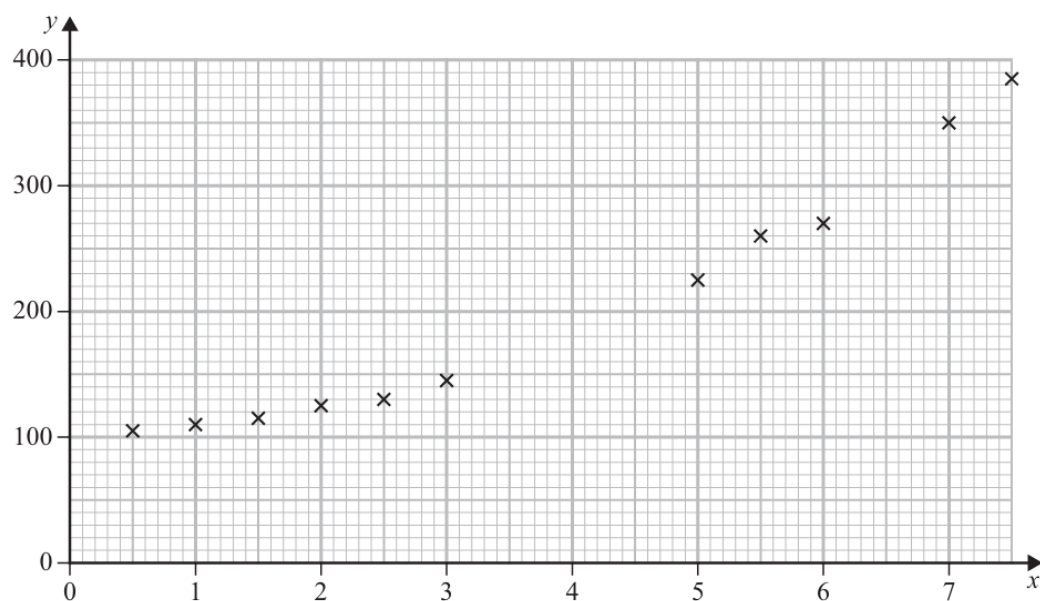


Figure 1

- (a) Describe the correlation between x and y

Positive

(1)

The equation of the regression line of y on x for these data is $y = 52.4 + 39.6x$

- (b) Give an interpretation of the gradient of this line.

average velocity of 39.6 m/s.

(1)

Nat drew the regression line on the scatter diagram to give Figure 2

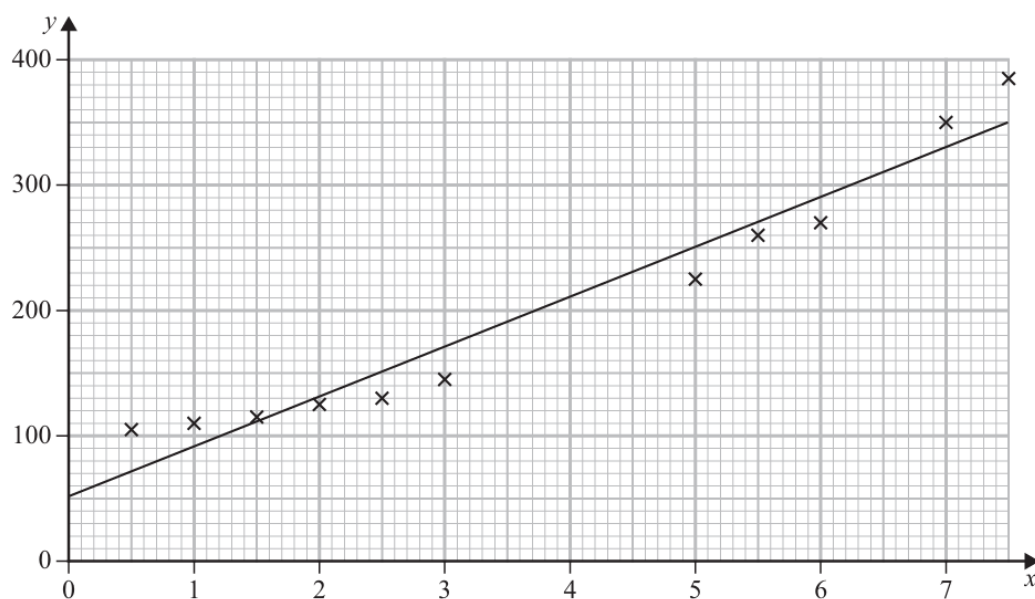


Figure 2

