Hypothesis Testing for correlation

Student	English Exam Mark	Maths Exam Mark
	Mean 60	Mean 70
	S.D. 5	S.D. 10
1	63.90	70.13
2	55.24	65.99
3	58.80	80.18
4	59.65	57.16
5	66.44	72.76
6	59.53	79.82
7	57.43	71.48
8	58.33	60.56
9	67.43	69.56
10	63.11	87.13
r=	0.219	

Student	English Exam Mark	Maths Exam Mark
	Mean 60	Mean 70
	S.D. 5	S.D. 10
1	60.22	74.64
2	62.25	79.15
3	61.30	75.29
4	60.61	71.35
5	55.31	74.05
6	57.13	89.73
7	57.16	70.41
8	58.96	60.31
9	56.30	71.95
10	63.23	69.95
r=	-0.094	

Suppose we use a spreadsheet to randomly generate maths marks for students, and separately generate random English marks.

(This Excel demo accompanies this file – you can press F9 in Excel to generate a new set of random data)

What is the **observed** PMCC between Maths and English marks in this first set of data?

But what is the true underlying PMCC between Maths and English?

How to carry out the hypothesis test

Student	English Exam Mark	Maths Exam Mark
	Mean 60	Mean 70
	S.D. 5	S.D. 10
1	63.90	70.13
2	55.24	65.99
3	58.80	80.18
4	59.65	57.16
5	66.44	72.76
6	59.53	79.82
7	57.43	71.48
8	58.33	60.56
9	67.43	69.56
10	63.11	87.13
r=	0.219	

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 Level 0.10	0.05	0.025	0.01	0.005	Sample Level	Spearman's Coefficient 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

Let's carry out a hypothesis test on whether there is positive correlation between English and Maths marks, at 10% significance level:

H_0:

H_1:

Sample size

Critical value for 10% significance level:

These values give the minimum value of r required to reject the null hypothesis, i.e. the amount of correlation that would be considered significant.

Two-tailed test

In the previous example we hypothesised that English/Maths marks were positively correlated. But we could also test whether there was **any** correlation, i.e. positive **or** negative.

[Textbook] A scientist takes 30 observations of the masses of two reactants in an experiment. She calculates a product moment correlation coefficient of $r = -0.45$.

The scientist believes there is no correlation between the masses of the two reactants. Test at the 10% level of significance, the scientist's claim, stating your hypotheses clearly.

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
0.2992	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
0.1678	0.2144	0.2542	0.2997	0.3301	60

H_0:

H_1:

Sample size $=$

Critical value at significance:

Test Your Understanding

[Textbook] The table from the large data set shows the daily maximum gust, x kn, and the daily maximum relative humidity, y%, in Leeming for a sample of eight days in May 2015.

x	31	28	38	37	18	17	21	29
y	99	94	87	80	80	89	84	86

a. Find the product moment correlation coefficient for this data.

b. Test, at the 10% level of significance, whether there is evidence of a positive correlation between daily maximum gust and daily maximum relative humidity. State your hypotheses clearly.