Hypothesis tests for correlation coefficients

Starter

1. A trading standards officer finds that the mean lifetime of a random sample of 50 long lasting light bulbs manufactured by En-Lightened Ltd is 1700 hours rather than the 1800 hours claimed by the company. The light bulbs are known to have a standard deviation of 125 hours. The officer carries out a hypothesis test at the $5\ \%$ level believing that light bulbs last shorter than the company claims. What conclusion does the officer reach?

Notes

At AS level, bivariate data (i.e. data with two variable) was plotted on a scatter diagram and a line of best fit was drawn. The closer the points are to the line the greater the correlation between the variables. While this gives us a visual understanding of the correlation between the variables, it would be better if we could represent it by a number.

Along came the English mathematician Karl Pearson (1857-1936), who devised ***Pearson's product moment correlation coefficient***, r.

Product moment correlation coefficient

The product moment correlation coefficient, r, is used with ***bivariate*** data as a measure of ***how closely related*** two variables are ***when it is evident there could be a linear relationship***.

Statisticians don't bother calculating the PMCC when it is clear the relationship is not linear.

Range of PMCC, r

The range of values of the PMCC is $-1 \leq r \leq 1$.
$r = 1$ means the points lie exactly along a straight line with a positive gradient
$r = -1$ means the points lie exactly along a straight line with a negative gradient
So ***r is a measure of how closely the points*** on the scatter graph ***are to the line of best fit***.

Hypothesis testing of the PMCC value

While a PMCC value close to 1 or -1 indicates strong correlation, we can carry out a hypothesis test in order to get a better idea about other values.

Success criteria — how to carry out a significance test for the PMCC value

In exam questions, the PMCC, or $r-$, value will be provided.

The null hypothesis is that the two variables are not correlated. Therefore, if we reject, we are saying there is evidence to suggest that there is a correlation.

1. State the null hypothesis: it is ***always*** $H_0 : \rho = 0$ where ρ is the population correlation coefficient
2. Decide whether you will carry out a $1-$tailed or $2-$tailed test (it is usually given in the question).
3. State the alternative hypothesis, H_1:
   - $H_1 : \rho \neq 0$ — two-tailed test
   - $H_1 : \rho > 0$ — one-tailed test, it is believed there could be a positive correlation
   - $H_1 : \rho < 0$ — one-tailed test, it is believed there could be a negative correlation
4. Decide on the level of significance (it is usually given in the question)
5. Use the tables to find the critical value based on the number of values, n, and the significance level
6. Compare your $r-$value to the critical value:

- If $\left| r - \text{value} \right| >$ critical value then reject $H_0 \Rightarrow$ there is evidence to suggest there is a correlation between the two variables
- If $\left| r - \text{value} \right| <$ critical value then do not reject $H_0 \Rightarrow$ there is no evidence to suggest there is a correlation between the two variables

The table of critical values for the product-moment correlation coefficient, r, is on p545 of the textbook.

N.B. The level of significance is the level of doubt that the statistician is willing to allow. Therefore, data must have a greater correlation to rejects H_0 at the $1\,\%$ level compared to the $5\,\%$ level.

E.g. 1 Carry out a one-tailed hypothesis test for these PMCC values at the given level of significance, where n is the sample size:
(a) $r = 0.465, n = 18$ at the $5\,\%$
(b) $r = 0.31, n = 45$ at the $1\,\%$
Write down the null and alternative hypotheses each time.

Working: (a) $H_0 : \rho = 0$
$H_1 : \rho > 0$
From tables, for a ***one-tailed test*** at the $5\,\%$ level where $n = 18$, the critical value is 0.4000.
Since $|r| = 0.465 > 0.4000$, reject H_0
There is sufficient evidence to suggest that there is positive correlation between the variables.

E.g. 2 Carry out a one-tailed hypothesis test for these PMCC values at the given level of significance, where n is the number of paired bivariate data:
(a) $r = -0.37, n = 23$ at the $2.5\,\%$
(b) $r = -0.89, n = 6$ at the $5\,\%$
Write down the null and alternative hypotheses each time.

Working: (a) $H_0 : \rho = 0$
$H_1 : \rho < 0$
From tables, for a ***one-tailed test***, at the $2.5\,\%$ level where $n = 23$, the critical value is 0.4132.
Since $|r| = 0.37 < 0.4132$, do not reject H_0
There is insufficient evidence to suggest that there is negative correlation between the variables.

E.g. 3 Carry out a two-tailed hypothesis test for these PMCC values at the given level of significance, where n is the sample size:
(a) $r = 0.751, n = 8$ at the $1\,\%$
(b) $r = -0.89, n = 11$ at the $5\,\%$
Write down the null and alternative hypotheses each time.

Working: (a) $H_0 : \rho = 0$
$H_1 : \rho \neq 0$
From tables, for a ***two-tailed test*** at the $1\,\%$ level where $n = 8$, the critical value is 0.8343.
Since $|r| = 0.751 < 0.8343$, do not reject H_0
There is insufficient evidence to suggest that there is correlation between the variables.

E.g. 4 A farming cooperative plotted mass of wheat crop yield against rainfall for 9 different areas of the country. The correlation coefficient for the data was 0.659.

(a) Test at the $5\,\%$ level whether rainfall improves wheat crop. State the null and alternative hypotheses clearly.

(b) Give a reason why the a hypothesis test may not be suitable for the data value.

Critical values of Pearson's product-moment correlation coefficient

1-tail test	5%	2.5%	1%	0.5%
2-tail test	10%	5%	2%	1%
n				
1	-	-	-	-
2	-	-	-	-
3	0.9877	0.9969	0.9995	0.9999
4	0.9000	0.9500	0.9800	0.9900
5	0.8054	0.8783	0.9343	0.9587
6	0.7293	0.8114	0.8822	0.9172
7	0.6694	0.7545	0.8329	0.8745
8	0.6215	0.7067	0.7887	0.8343
9	0.5822	0.6664	0.7498	0.7977
10	0.5494	0.6319	0.7155	0.7646
11	0.5214	0.6021	0.6851	0.7348
12	0.4973	0.5760	0.6581	0.7079
13	0.4762	0.5529	0.6339	0.6835
14	0.4575	0.5324	0.6120	0.6614
15	0.4409	0.5140	0.5923	0.6411
16	0.4259	0.4973	0.5742	0.6226
17	0.4124	0.4821	0.5577	0.6055
18	0.4000	0.4683	0.5425	0.5897
19	0.3887	0.4555	0.5285	0.5751
20	0.3783	0.4438	0.5155	0.5614
21	0.3687	0.4329	0.5034	0.5487
22	0.3598	0.4227	0.4921	0.5368
23	0.3515	0.4132	0.4815	0.5256
24	0.3438	0.4044	0.4716	0.5151
25	0.3365	0.3961	0.4622	0.5052
26	0.3297	0.3882	0.4534	0.4958
27	0.3233	0.3809	0.4451	0.4869
28	0.3172	0.3739	0.4372	0.4785
29	0.3115	0.3673	0.4297	0.4705
30	0.3061	0.3610	0.4226	0.4629

Video: Hypothesis testing for zero correlation

Video: Correlation hypothesis testing

Solutions to Starter and E.g.s

Exercise

p410 18C Qu 1i, 2i, 3-7, (8-9 red)

Summary

Product moment correlation coefficient:

The product moment correlation coefficient, r, is used with ***bivariate*** data as a measure of ***how closely related*** two variables are ***when it is evident there could be a linear relationship***.

The range of values of the PMCC is $-1 \leq r \leq 1$.
$r = 1$ means the points lie exactly along a straight line with a positive gradient
$r = -1$ means the points lie exactly along a straight line with a negative gradient
r is a measure of how closely the points on the scatter graph ***are to the line of best fit***.

Success criteria — how to carry out a significance test for the PMCC value:
The null hypothesis is that the two variables are not correlated. Therefore, if we reject, we are saying there is evidence to suggest that there is a correlation.

1. State the null hypothesis: it is ***always*** $H_0 : \rho = 0$ where ρ is the population correlation coefficient
2. Decide whether you will carry out a $1-$tailed or $2-$tailed test (it is usually given in the question).
3. State the alternative hypothesis, H_1:
   - $H_1 : \rho \neq 0$ — two-tailed test
   - $H_1 : \rho > 0$ — one-tailed test, it is believed there could be a positive correlation
   - $H_1 : \rho < 0$ — one-tailed test, it is believed there could be a negative correlation
4. Decide on the level of significance (it is usually given in the question)
5. Use the tables to find the critical value based on the number of values, n, and the significance level
6. Compare your $r-$value to the critical value:
   - If $\left| r - \text{value} \right| >$ critical value then reject $H_0 \Rightarrow$ there is evidence to suggest there is a correlation between the two variables
   - If $\left| r - \text{value} \right| <$ critical value then do not reject $H_0 \Rightarrow$ there is no evidence to suggest there is a correlation between the two variables

The table of critical values for the product-moment correlation coefficient, r, is on p545 of the textbook.

The level of significance is the level of doubt that the statistician is willing to allow. Therefore, data must have a greater correlation to reject H_0 at the $1\,\%$ level compared to the $5\,\%$ level.