Visit http://www.mathsmadeeasy.co.uk/ for more fantastic resources.

OCR

# A Level

# A Level Mathematics

## Pearson's Correlation Coefficient (Answers)

Name:

MME

Mathsmadeeasy.co.uk

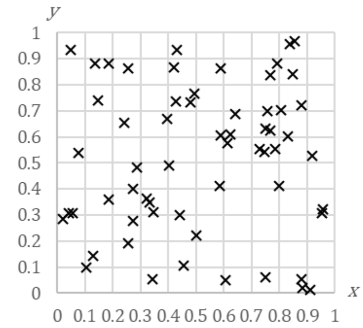
Total Marks:

D1- Pearson's Correlation Coefficient- Answers

OCR

**1) Estimate the correlation coefficient of the data shown in each of the following graphs.**

[1 mark for each correct answer in acceptable range- 3 max]



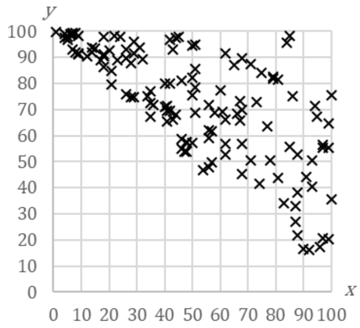

Has no observable correlation.

Acceptable range:

$$-0.1 \leq \rho \leq 0.1$$

Actual

$$\rho = -0.08$$



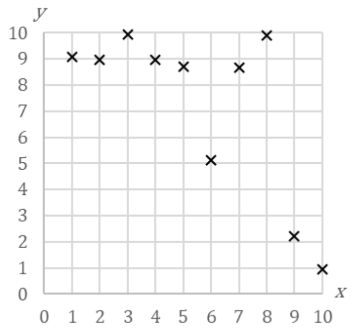

Has negative correlation, with some spread.

Acceptable range:

$$-0.8 \leq \rho \leq -0.6$$

Actual:

$$\rho = -0.71$$




Negative correlation, with spread, few observations.

Acceptable range:

$$-0.75 \leq \rho \leq -0.55$$

Actual:

$$\rho = -0.69$$

2) **For each of the following determine whether $H_0: \rho = 0$ can be rejected or accepted and at what level of significance- either 5, 1, 0.1%.**

| | $\rho$ | $n$ | P-value |
|---|---|---|---|
| i) | 0.73 | - | 0.00169 |
| ii) | 0.86 | - | 0.0495 |
| iii) | 0.977 | 12 | - |

[1 mark]

i) $H_0$ can be rejected in favour of the alternate hypothesis at 5%,1% and 0,1% level of significance. As p < 0.01%.

[1 mark]

ii) $H_0$ can be rejected in favour of the alternate hypothesis at 5% level of significance. As p < 5%.

[1 mark]

Calculate the t-critical value given by

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

[1 mark]

$$t = 0.977\sqrt{\frac{12-2}{1-0.977^2}}$$

$$t = 14.48862964$$

[1 mark]

$$v - 2\ degrees\ of\ freedom = 10$$

[1 mark for each correct critical value- 3 max]

If $obs\ t > \ crit\ t$ then reject $H_0$.

$$t_{10}(0.025) = 2.228138852$$

$$t_{10}(0.005) = 3.169272673$$

$$t_{10}(0.00005) = \ 4.586893859$$

**3) The results of a machine that learns from its mistakes are shown in the table below.**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Number of Experiments** | **10** | **20** | **30** | **40** | **50** | **60** | **70** | **80** | **90** | **100** |
| **Accuracy** | **0** | **3** | **9** | **15** | **23** | **37** | **52** | **94** | **99** | **100** |

[2]

[2]

**i) Calculate Pearson's correlation coefficient.**

Let $x$ be Number of Experiments and $y$, Accuracy.

[1 mark]

$$\Sigma x = 550$$

$$\bar{x} = 55$$

$$\Sigma(x - \bar{x})^2 = 8250$$

[1 mark]

$$\Sigma y = 432$$

$$\bar{y} = 432.2$$

$$\Sigma(y - \bar{y})^2 = 14891.6$$

[1 mark]

$$r = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\sqrt{\Sigma(x - \bar{x})^2 \, \Sigma(y - \bar{y})^2}}$$

[1 mark]

$$r = \frac{10610}{\sqrt{(8250)(14891.6)}}$$

$$r = 0.9572$$

**ii) Write null and alternate hypotheses regarding the significant of the calculated coefficient.**

[1 mark for each hypothesis – 2 max]

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

$$\therefore two - tailed\ test$$

**iii) Carry out a t-test at the 5% significance level. Use this to reject or accept the null hypothesis.**

[1 mark]

Calculate the t-critical value given by

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

[1 mark]

$$t = 0.9572\sqrt{\frac{10-2}{1-0.9572^2}}$$

$$t = 9.354$$

[1 mark]

$$v - 2 \; degrees \; of \; freedom$$

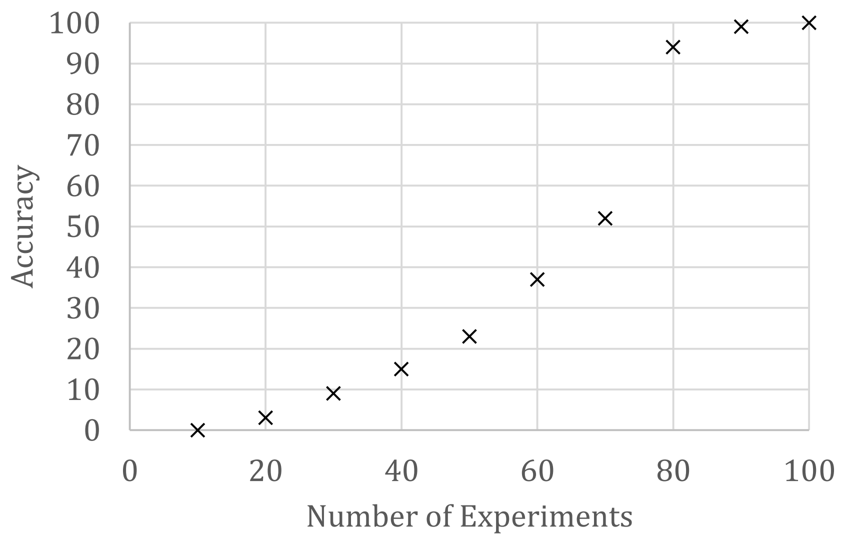
$$\therefore t \; crit = t_8(0.05) = 2.306004135$$

[1 mark]

As $obs \; t \gg t \; crit$ we can reject the null hypothesis in favour of the alternate hypothesis. i.e. there is strong evidence to suggest a correlation.

**iv) Plot a suitable graph of the data.**

[1 mark- a scatter plot]




**v) Without calculation state the effect, with a reason, that removing the first two pairs (10,0) and (20,3) would have.**

[1 mark]

It would increase the value of the correlation coefficient as the points would be "straighter".

**vi) Assuming the Accuracy increases linearly between intervals calculate the Accuracy after 66 experiments.**

[1 mark]

Using $difference\ in\ y - difference\ in\ x$ we can gain an understanding that in the interval 60 to 70, for every experiment the accuracy increases by 1.5.

[1 mark]

Therefore after 66 experiments the Accuracy is 46.

**vii) Explain why $\boldsymbol{Number\ of\ Experiments}$ is the independent variable.**

[1 mark for any of following – 1 max]

- Not random
- Not affected by y
- Regular intervals
- Not being measured
- Not dependent on anything