Pearson's Product Moment Correlation Coefficient

Starter

1. **(Review of last lesson)** The number of accidents per week at a factory is a Poisson random variable with parameter 2.

   (a) Find the probability that in any week chosen at random exactly 1 accident occurs.
   (b) The factory is observed for 100 weeks. Determine the expected number of weeks, to 1 d.p., in which 5 or more accidents occur.

Working: (a) $X \sim \text{Po}(2)$

Formula: $P(X = 1) = \dfrac{e^{-2} \times 2^1}{1!} = 0.2707$

(b) $X \sim \text{Po}(2)$

$$P(X \geq 5) = 1 - P(X \leq 4) = 0.05265$$

Assume the 100 weeks follows a binomial distribution, where

$Y \sim \text{B}(100, 0.5265)$

The expected number of weeks is:

$$E(Y) = np = 100 \times 0.5265 = 5.3$$

Video : **Calculating mean and variance using a Classwiz calculator**

Enter data >> AC >> OPTN >> 2

2. For the following data, use the 6: Statistics function on your calculator to find:

   (a) the mean for X and the mean for Y
   (b) the standard deviation of X and the standard deviation of Y.

X	2130	2424	2328	2394	2280	2232	2082	2274
Y	2244	2430	2376	2388	2226	2166	2202	2208

Working: (a) $\mu_X = 2268; \mu_Y = 2280$

(b) $\sigma_X = 111.2; \sigma_Y = 94.77$

2. Eight runners ran the same 10 km route twice — once in the dry and once when it was wet. Their times, in minutes, are below:

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**
Dry, D	35.5	40.4	38.8	39.9	38.0	37.2	34.7	37.9
Wet, W	37.4	40.5	39.6	39.8	37.1	36.1	36.7	36.8

Use the 6: Statistics function on your calculator to find:

(a) the mean time for the dry and the wet.
(b) the standard deviation for the dry and the wet.

Working: (a) $\mu_{dry} = 37.8, \mu_{wet} = 38.0$

(b) $\mu_{dry} = 1.853, \mu_{wet} = 1.580$

E.g. 1 Without using the special function on your calculator, find the PMCC for $\sum x_i = 22.09$, $\sum y_i = 49.7$, $\sum x_i^2 = 45.04$, $\sum y_i^2 = 244.83$, $\sum x_i y_i = 97.778$ and $n = 12$.

Working:

$$S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} = 97.778 - \frac{22.09 \times 49.7}{12} = 6.2886 \text{ (4 d.p.)}$$

$$S_{xx} = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n} = 45.04 - \frac{22.09^2}{12} = 4.3759 \text{ (4 d.p.)}$$

$$S_{yy} = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n} = 244.83 - \frac{49.7^2}{12} = 38.989 \text{ (4 d.p.)}$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \frac{6.2886}{\sqrt{4.3759 \times 38.989}} = 0.481 \text{ (3 d.p.)}$$

E.g. 2 In the study of a city, the population density p (people/hectare) and the distance from the city centre d (km) was investigated by picking a number of random areas with the following results.

Area	A	B	C	D	E	F	G	H	I	J
Distance, d (km)	0.6	3.8	2.4	3.0	2.0	1.5	1.8	3.4	4.0	0.9
Population density, p (people/hectare)	50	22	14	20	33	47	25	8	16	38

Calculate the PMCC and comment on your findings.

Working: $r = -0.823$, there is strong negative correlation between distance from the centre of the city centre and the population density

Video: Pearson's PMCC
Video: PMCC on calculators
Video : Calculating mean and variance using a Classwiz calculator

Solutions to Starter and E.g.s

Exercise
p75 5A Qu 1i, 2i