Stats Jan 05.

1\. i Set A has the largest value of PMCC as it has good correlation so a PMCC value of almost 1.

ii Set C has the smallest value as it has no linear relationship so PMCC will be close to 0.

2 i LQ - 6 hours
Med - 8 hours
UQ - 24 hours

ii Because the data is skewed.

iii Advantage - All data is shown in full
Disadvantage - Harder to spot IQR & Med

3\.

~~A~~	Rank 1	Rank 2	d	d^2
A	6	7	1	1
B	5	6	1	1
C	4	5	1	1
D	7	4	3	9
E	2	3	1	1
F	3	2	1	1
G	1	1	0	0
				14

i $r_s = 1 - \frac{6\sum d^2}{7(7^2-1)} = 1 - \frac{84}{336}$
$= \underline{\underline{0.75}}$

ii There is a strong correlation between their scores - They are in good agreement.

4 i $k = 1 - \left(\frac{1}{4} + \frac{1}{5} + \frac{2}{5} + \frac{1}{10}\right)$
$k = \frac{1}{20}$.

ii $E(x) = \sum x_i p_i = (-2 \times \frac{1}{4}) + (-1 \times \frac{1}{5}) + (0 \times \frac{1}{20}) + (1 \times \frac{2}{5}) + (2 \times \frac{1}{10})$

$E(x) = -\frac{1}{10}$.

$Var(x) = \sum x_i^2 p_i = (-2^2 \times \frac{1}{4}) + (-1^2 \times \frac{1}{5}) + (0^2 \times \frac{1}{20}) + (1^2 \times \frac{2}{5}) + (2^2 \times \frac{1}{10})$.

$Var(x) = 2. - \mu^2 = 2 - \frac{1}{100} = \underline{\underline{1.99}}$

5 i $X \sim Geo(\frac{1}{20})$.

a $P(X=6) = \left(\frac{19}{20}\right)^5 \times \frac{1}{20}$
$= 0.0387$.

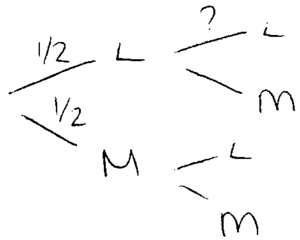
b $P(X>10) = \left(\frac{19}{20}\right)^{10}$
$= 0.599$.

ii $E(x) = \frac{1}{p} = \underline{\underline{20}}$.

6\. X = prob that Louise wins first 2 sets.

i $X \sim B(5, \frac{3}{8})$.
$P(X=2) = {}^5C_2 \times \left(\frac{3}{8}\right)^2 \times \left(\frac{5}{8}\right)^3$
$= 0.343$.

ii



$= 3/8$.

So $\frac{1}{2} \times ? = 3/8$.
P(L wins 2nd | wins 1st)
$= 3/4$

iii P(M wins 1st 2) $= 1/3$.
$\frac{1}{2} \times ? = \frac{1}{3}$. $? = 2/3$.

P(M wins 2nd | wins 1st) $= 2/3$.

7 i Boxes must be independant of each other.
Probabilities must remain fixed.

ii a $P(Y=0)$
$Y \sim B(8, 1/10)$

$P(Y=0) = {}^8C_0 \, 9/10^8$
$= 0.43$

b. $P(Y \geq 2) = 1 - (P(Y = 0 \text{ or } 1)$

$P(Y=1) = {}^8C_1 \times 9/10^7 \times 1/10$
$= 0.38$

$1 - (0.43 + 0.38)$
$P(Y \geq 2) = 0.187$

iii
$P(W=0 \wedge T \geq 2)$ or $P(W \geq 2 \wedge T = 0)$.
$= 0.43 \times 0.187 \times 2 = 0.161$.

8 i 8Q's $P(\text{BG or GB})$.
$= \frac{7! \times 2}{8!} = 1/4$.

ii 4 × 7 marks (1G) 4 × 8 marks (1B).

$\underbrace{\_\ \_\ \_}_{3!}$ G B $\underbrace{\_\ \_\ \_}_{3!}$ 3! × 3! ways.

4! × 4! ways altogether

$P(\ ) = \frac{3!\,3!}{4!\,4!} = 1/16$.

iii
_ _ _ G _ _ B _ or
_ _ _ G _ _ _ B or
_ _ G _ _ B _ _ or
_ _ G _ _ _ B _ or
_ _ G _ _ _ _ B or
_ G _ _ B _ _ _ or
_ G _ _ _ B _ _ or
_ G _ _ _ _ B _ or
_ G _ _ _ _ _ B or
G _ _ _ B _ _ _ or
G _ _ _ _ B _ _ or
G _ _ _ _ _ B _ or
G _ _ _ _ _ _ B.

If not separated by 2.

_ _ _ G B _ _ _ or $= 1/16$
_ _ _ G _ B _ _ or $= 1/16$
_ _ G _ B _ _ _ $= 1/16$.

$P(\text{not seperated by } 2) = 3/16$.

$P(\text{seperated by 2 or more}) = 1 - 3/16$
$= 13/16$.

9. i y on x $\Rightarrow$ $y = a + bx$.

so gradient $= b = \frac{S_{xy}}{S_{xx}}$

$S_{xy} = \sum xy - \frac{\sum x \sum y}{5} = 264 - \frac{90 \times 15}{5} = -6$.

$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{5} = 1720 - \frac{90^2}{5} = 100$

$\frac{S_{xy}}{S_{xx}} = \frac{-6}{100} = -0.06$.

$a = \bar{y} - b\bar{x}$ $\quad a = \frac{15}{5} - (-0.06)\frac{90}{5}$

$a = 4.08$

$y = 4.08 - 0.06x$

ii $\quad x = 20$ nearest whole $\Rightarrow$ $19.5 \leq x < 20.5$

If $x = 19.5$ $\quad y = 2.91$
$x = 20.5$ $\quad y = 2.85$

~~$2.91 \leq y \leq 2.85$~~ $\quad 2.91 - 2.85 = \underline{0.06}$

iii $\quad e_4 = a + b(23) - (3.3)$
$= 4.08 + -0.06(23) - (3.3)$
$= -0.6$

$e_5 = a + b(23) - (2.2)$
$= 4.08 + (-0.06)(23) - (2.2)$
$= 0.5$

iv $\quad 0.6^2 + (-0.7)^2 + 0.2^2 + (-0.6)^2 + 0.5^2$
$= 1.5$

This is the minimised distance from the regression line.

V. $\quad$ Mean $\Rightarrow \frac{0.6 - 0.7 + 0.2 - 0.6 + 0.5}{5}$

$= 0$

Variance $\rightarrow \frac{\sum x_i^2}{5} - 0^2$

$= 0.6^2 + 0.7^2 + 0.2^2 + 0.6^2 + 0.5^2$

$= \frac{1.5}{5} = \underline{0.3}$