2006-Jan

1 Jenny and John are each allowed two attempts to pass an examination.

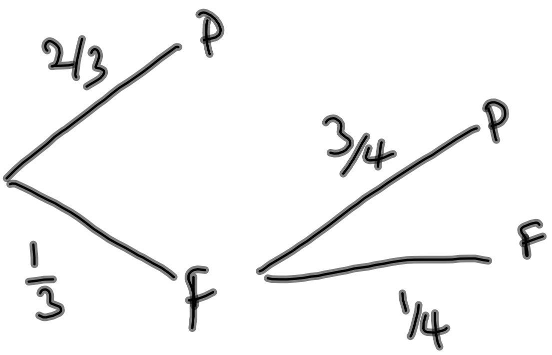
(i) Jenny estimates that her chances of success are as follows.

- The probability that she will pass on her first attempt is $\frac{2}{3}$.
- If she fails on her first attempt, the probability that she will pass on her second attempt is $\frac{3}{4}$.

Calculate the probability that Jenny will pass. [3]

$P(P) = 2/3 = \frac{8}{12}$

$P(F, P) = \frac{1}{3} \times \frac{3}{4} = \frac{3}{12}$



$P(F, P) = \frac{1}{3} \times \frac{1}{4} = \frac{1}{12}$

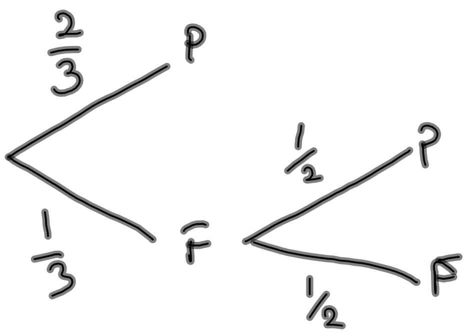
$P(\text{Pass}) = \frac{8}{12} + \frac{3}{12} = \frac{11}{12}$ ✓

or 0.917 (3sf)

(ii) John estimates that his chances of success are as follows.

- The probability that he will pass on his first attempt is $\frac{2}{3}$.
- Overall, the probability that he will pass is $\frac{5}{6}$.

Calculate the probability that if John fails on his first attempt, he will pass on his second attempt. [3]



$2/3$

$1/6$

$\frac{1}{6}$

If P(Pass 1st attempt) $= 2/3$

then P(Fail 1st attempt) $= 1/3$

P(Fail overall) $= 1 - 5/6 = \frac{1}{6}$

P(Fail overall) $= \frac{1}{3} \times x = \frac{1}{6}$

so P(Fail 2nd time) $= \frac{1}{2}$

$\therefore$ P(Pass 2nd time) $= 1 - \frac{1}{2} = \frac{1}{2}$

P(Pass overall) $= \frac{2}{3} +$ P(F, P) $= \frac{5}{6}$

$= \frac{4}{6} + \left(\frac{1}{3} \times \frac{1}{2}\right) = \frac{5}{6}$

P(Passing 2nd time) $= \frac{1}{2}$ ✓

* Find the solution by working out the completed probability tree.

2006-Jan

2 For each of 50 plants, the height, h cm, was measured and the value of $(h - 100)$ was recorded. The mean and standard deviation of $(h - 100)$ were found to be 24.5 and 4.8 respectively.

(i) Write down the mean and standard deviation of h. [2]

mean $= 24.5 + 100 = 124.5$ ✓

sd $= 4.8$ ✓

The mean and standard deviation of the heights of another 100 plants were found to be 123.0 cm and 5.1 cm respectively.

(ii) Describe briefly how the heights of the second group of plants compare with the first. [2]

The mean ~~of~~ of the 2nd batch is 1.5cm less ✓ generally smaller

and the s.d. 0.3cm more ✓ more dispersed

(iii) Calculate the mean height of all 150 plants. [2]

$$\sum x = (50 \times 124.5) + (100 \times 123)$$

$$= 6225 + 12300$$

$$= 18525$$

$$\bar{x} = \frac{\sum x}{n} = \frac{18525}{150} = 123.5$$ ✓

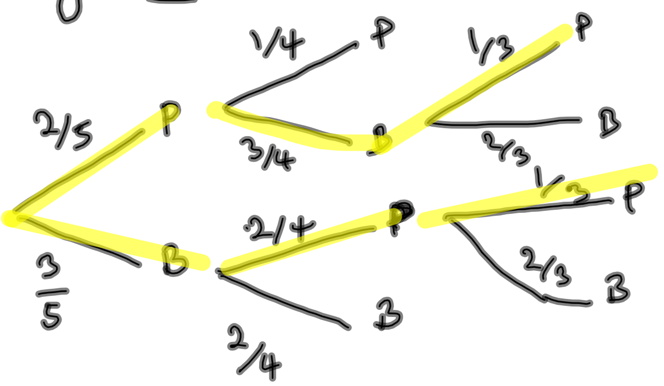
2006-Jan

3 In Mr Kendall's cupboard there are 3 tins of baked beans and 2 tins of pineapple. Unfortunately his daughter has removed all the labels for a school project and so the tins are identical in appearance. Mr Kendall wishes to use both tins of pineapple for a fruit salad. He opens tins at random until he has opened the two tins of pineapples.

Let X be the number of tins that Mr Kendall opens.

(i) Show that $P(X = 3) = \frac{1}{5}$. [4]

Probability is not constant



$\frac{2}{5} \times \frac{3}{4} \times \frac{1}{3} = \frac{6}{60}$

$\frac{3}{5} \times \frac{2}{4} \times \frac{1}{3} = \frac{6}{60}$

must be either B, P, P or P, B, P for $X = 3$

$P(X=3) = P(B,P,P) + P(P,B,P)$

$= \frac{6}{60} + \frac{6}{60} = \frac{12}{60} = \frac{1}{5}$ ✓

(ii) The probability distribution of X is given in the table below.

x	2	3	4	5
$P(X = x)$	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{3}{10}$	$\frac{2}{5}$

Find $E(X)$ and $\text{Var}(X)$. [5]

xp	$\frac{2}{10}$	$\frac{6}{10}$	$\frac{12}{10}$	$\frac{20}{10}$	$\sum xp = \frac{40}{10} = 4$
x^2	4	9	16	25	
x^2p	$\frac{4}{10}$	$\frac{18}{10}$	$\frac{48}{10}$	$\frac{100}{10}$	$\sum x^2p = \frac{170}{10} = 17$

$\mu = E(x) = \sum xp = \underline{\underline{4}}$ ✓

$Var(x) = \sum x^2 p - \mu^2$

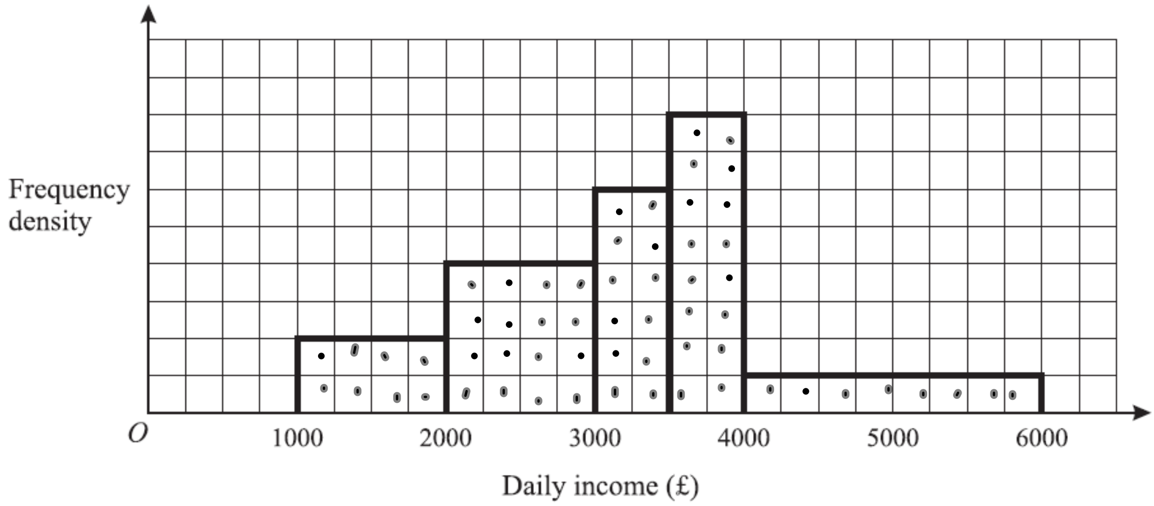
$= 17 - 16$

$= \underline{\underline{1}}$ ✓

2006-Jan

4 Each day, the Research Department of a retail firm records the firm's daily income, to be used for statistical analysis. The results are summarised by recording the number of days on which the daily income is within certain ranges.

(i)



The histogram shows the results for 300 days. By considering the total area of the histogram,

(a) find the number of days on which the daily income was between £4000 and £6000, [4]

All the squares represent 300 days

There are 60 squares ✓

Each square represents $\frac{300}{60} = 5$ days ✓

There are 8 squares between £4000 and £6000

This represents $8 \times 5 = 40$ days ✓

(b) calculate an estimate of the number of days on which the daily income was between £2700 and £3200. [3]

To calculate this estimate split the range either side of the £3000 category boundary. You need two estimates. One for £2700 → £3000 and one for £3000 → £3200.

RANGE	FREQUENCY
£2000 – £3000	$16 \times 5 = 80$
£3000 – £3500	$12 \times 5 = 60$

ESTIMATE: 2700 → 3000 $\frac{3}{10} \times 80 = 24$

3000 → 3200 $\frac{2}{5} \times 60 = 24$

2700 → 3200 $24 + 24 = 48$ ✓

(ii) The Research Department offers to provide any of the following statistical diagrams: histogram, frequency polygon, box-and-whisker plot, cumulative frequency graph, stem-and-leaf diagram and pie chart.

Which one of these statistical diagrams would most easily enable managers to

(a) read off the median and quartile values of the daily income, [1]

A box plot ✓

(b) find the range of the top 10% of values of the daily income? [1]

A cumulative frequency diagram ✓

2006-Jan

5 Andrea practises shots at goal. For each shot the probability of her scoring a goal is $\frac{2}{5}$. Each shot is independent of other shots.

(i) Find the probability that she scores her first goal

(a) on her 5th shot, [2]

This question is asking about the NUMBER OF TRIALS to the FIRST SUCCESS so we need to use a GEOMETRIC distribution

$$X \sim Geo\left(\frac{2}{5}\right)$$

$$P(X=5) = \left(\frac{3}{5}\right)^4 \times \frac{2}{5} = \frac{81}{625} \times \frac{2}{5}$$

$$= \frac{162}{3125}$$ ✓ or 0.0518 (3sf)

(b) before her 5th shot. [3]

$$P(X<5) = 1 - P(X>4)$$

$$= 1 - \left(\frac{3}{5}\right)^4$$

$$= 1 - \frac{81}{625}$$

$$= \frac{544}{625}$$ ✓ or 0.870 (3sf)

(ii) (a) Find the probability that she scores exactly 1 goal in her first 5 shots. [3]

Now we need the BINOMIAL distribution

$Y \sim B(5, \frac{2}{5})$

$P(Y=1) = \binom{5}{1} \times \frac{2}{5} \times \left(\frac{3}{5}\right)^4$

$= 5 \times \frac{2}{5} \times \frac{81}{625}$

$= \frac{5 \times 2 \times 81}{3125} = \frac{810}{3125} = 0.259$ (3sf)

(b) Hence find the probability that she scores her **second** goal on her 6th shot. [2]

To score the 2nd goal on the 6th shot she needs to have scored just one goal in the 1st 5 shots (which we worked out the probability for in the previous part)

$\frac{810}{3125} \times \frac{2}{5} = \frac{1620}{15625} = 0.10368$

$= 0.104$ (3sf)

2006-Jan

6 An examination paper consists of two parts. Section A contains questions A1, A2, A3 and A4. Section B contains questions B1, B2, B3, B4, B5, B6 and B7.

Candidates must choose three questions from section A and four questions from section B. The order in which they choose the questions does not matter.

(i) In how many ways can the seven questions be chosen? [3]

$$4C3 \times 7C4$$

$$= \frac{4!}{3!1!} \times \frac{7!}{4!3!} = 4 \times \frac{7 \times \cancel{6} \times 5}{\cancel{3 \times 2 \times 1}}$$

$$= \underline{\underline{140}} \checkmark$$

(ii) Assuming that all selections are equally likely, find the probability that a particular candidate chooses question A1 but does **not** choose question B1. [3]

$A1 \underline{\times} \underline{\times}$

No. of Combinations with A1

$$= 3C2 = \frac{3!}{2!1!} = 3$$

$$P(A1) = \frac{3}{4} \quad \left(\frac{3C2}{4C3}\right)$$

$B1 \underline{\times} \underline{\times} \underline{\times}$

$$6C3 = \frac{6!}{3!3!}$$

$$= \frac{\cancel{6} \times 5 \times 4}{\cancel{3 \times 2 \times 1}}$$

$$P(B1) = \frac{6C3}{7C4} = \frac{20}{35} = \frac{4}{7}$$

$$P(B1') = 1 - \frac{4}{7} = \frac{3}{7}$$

$$P(A1 \cap B1') = \frac{3}{4} \times \frac{3}{7} = \frac{9}{28} \checkmark \text{ or } 0.321 \text{ (3sf)}$$

(iii) Following a change of syllabus, the form of the examination remains the same except that candidates who choose question A1 are not allowed to choose question B1. In how many ways can the seven questions now be chosen? [3]

The simplest way to think about this is to consider the number of combinations that don't include A1 and the number that do include A1 separately.

section A w/o A1 × any section B (7C4)

1 × 35 = 35

section w. A1 × section B w/o B1 (7C4 - 6C3)
35 - 20

3 × 15 = 45

TOTAL POSSIBLE COMBINATIONS

35 + 45 = 80

2006-Jan

7 Past experience has shown that when seeds of a certain type are planted, on average 90% will germinate. A gardener plants 10 of these seeds in a tray and waits to see how many will germinate.

(i) Name an appropriate distribution with which to model the number of seeds that germinate, giving the value(s) of any parameters. State any assumption(s) needed for the model to be valid. [4]

Binomial $X \sim B(10, 0\cdot9)$ ✓

- the probability of each seed germinating must be constant ✓
- the probability of any seed germinating must be independent of the probability of any other seed germinating. ✓

(ii) Use your model to find the probability that fewer than 8 seeds germinate. [2]

$P(X<8) = P(X \leq 7)$

$= 0\cdot0702$ ✓ (from tables)

Later the gardener plants 20 trays of seeds, with 10 seeds in each tray.

(iii) Calculate the probability that there are at least 19 trays in each of which at least 8 seeds germinate. [4]

Here we need to set up a new BINOMIAL MODEL using our answer to 7(ii)

$$P(X \geq 8) = 1 - 0.0702$$
$$= 0.9298$$

$$Y \sim B(20, 0.9298)$$

$$P(Y \geq 19) = P(Y=19) + P(Y=20)$$

$$= (20 \times 0.9298^{19} \times 0.0702) + (1 \times 0.9298^{20})$$

$$= 0.58541646$$

$$= 0.585 \text{ (3sf)}$$

8 The table shows the population, x million, of each of nine countries in Western Europe together with the population, y million, of its capital city.

	Germany	United Kingdom	France	Italy	Spain	The Netherlands	Portugal	Austria	Switzerland
x	82.1	59.2	59.1	56.7	39.2	15.9	9.9	8.1	7.3
y	3.5	7.0	9.0	2.7	2.9	0.8	0.7	1.6	0.1

$[n = 9,\ \Sigma x = 337.5,\ \Sigma x^2 = 18\,959.11,\ \Sigma y = 28.3,\ \Sigma y^2 = 161.65,\ \Sigma xy = 1533.76.]$

(i) (a) Calculate Spearman's rank correlation coefficient, r_s. [5]

This is a RANK correlation so we have to RANK the data.

	G	UK	F	I	S	N	P	A	S
x	82.1	59.2	59.1	56.7	39.2	15.9	9.9	8.1	7.3
y	3.5	7.0	9.0	2.7	2.9	0.8	0.7	1.6	0.1
Rank x	1	2	3	4	5	6	7	8	9 ✓
Rank y	3 ✓	2 ✓	1 ✓	5 ✓	4 ✓	7 ✓	8 ✓	6 ✓	9 ✓
d	−2	0	2	−1	1	−1	−1	2	0
d^2	4	0	4	1	1	1	1	4	0

$$\Sigma d^2 = 4 + 4 + 1 + 1 + 1 + 1 + 4 = 16$$

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 16}{9 \times 80}$$

$$= \frac{13}{15} \checkmark = 0.867 \checkmark \ (3sf)$$

(b) Explain what your answer indicates about the populations of these countries and their capital cities. [1]

$r_s = 0.867$

This shows a strong correlation between the population of the country and the population of its capital city (r_s is close to 1)

Mark scheme BE EXPLICIT : larger countries tend to have larger capital cities in terms of population

(ii) Calculate the product moment correlation coefficient, r. [2]

We've been given the summary statistics so no need to calculate them!

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

$$S_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 1533.76 - \frac{337.5 \times 28.3}{9}$$

$$= 472.51$$

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 18959.11 - \frac{337.5^2}{9}$$

$$= 6302.86$$

$$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 161.65 - \frac{28.3^2}{9}$$

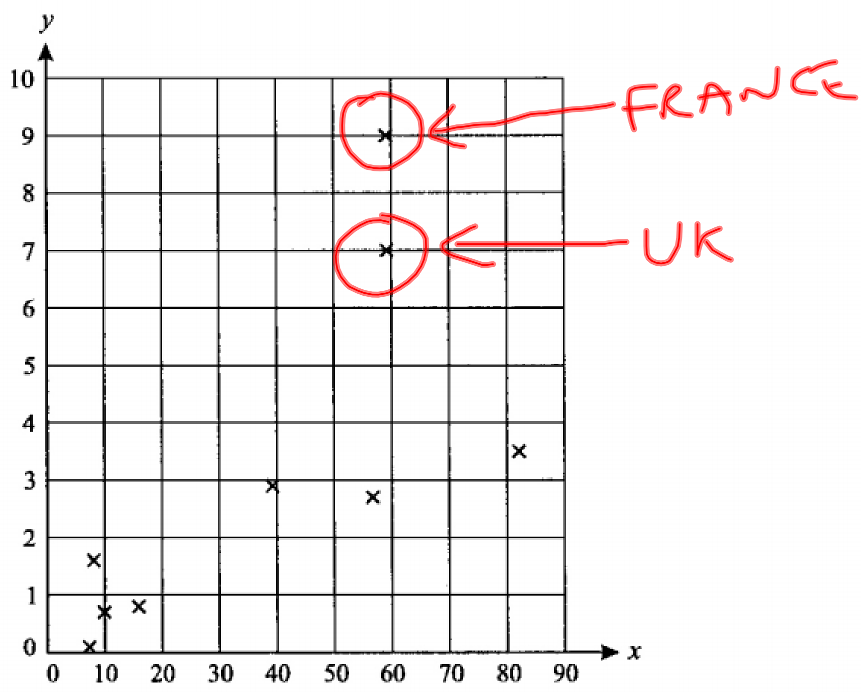
$$= 72.662 \checkmark$$

$$r = \frac{472.51}{\sqrt{6302.86 \times 72.662}}$$

$$= 0.6982139155$$

$$= \underline{\underline{0.698}} \text{ (3sf)} \checkmark$$

The data are illustrated in the scatter diagram.



(iii) By considering the diagram, state the effect on the value of the product moment correlation coefficient, r, if the data for France and the United Kingdom were removed from the calculation. [1]

If the UK and France are removed from the calculation of the product moment correlation coefficient then r will get bigger (closer to 1) because the remaining data points line up more closely to a straight line.

(iv) In a certain country in Africa, most people live in remote areas and hence the population of the country is unknown. However, the population of the capital city is known to be approximately 1 million. An official suggests that the population of this country could be estimated by using a regression line drawn on the above scatter diagram.

(a) State, with a reason, whether the regression line of y on x or the regression line of x on y would need to be used. [2]

We are trying to predict the population of the country from the population of the capital so x (country) is the DEPENDENT variable. We need to regress the DEPENDENT variable on the INDEPENDENT variable so x on y would need to be used.

(b) Comment on the reliability of such an estimate in this situation. [2]

In order to be reliable we would need to assume that the population structure of this African country is very similar to that of the European countries for which we have complete data. This is not necessarily the case so the estimate might not be reliable.