Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel Level 3 GCE

Centre Number | Candidate Number

Paper reference **8MA0/21**

Mathematics

Advanced Subsidiary

PAPER 21: Statistics

You must have:
Mathematical Formulae and Statistical Tables, calculator

Total Marks

Candidates may use any calculator allowed by Pearson regulations. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulae stored in them.

Instructions

- Use **black** ink or ball-point pen.
- If pencil is used for diagrams/sketches/graphs it must be dark (HB or B).
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions and ensure that your answers to parts of questions are clearly labelled.
- Answer the questions in the spaces provided
  – *there may be more space than you need.*
- You should show sufficient working to make your methods clear. Answers without working may not gain full credit.
- Values from statistical tables should be quoted in full. If a calculator is used instead of tables the value should be given to an equivalent degree of accuracy.
- Inexact answers should be given to three significant figures unless otherwise stated.

Information

- A booklet 'Mathematical Formulae and Statistical Tables' is provided.
- The total mark for this part of the examination is 30. There are 5 questions.
- The marks for **each** question are shown in brackets
  – *use this as a guide as to how much time to spend on each question.*

Advice

- Read each question carefully before you start to answer it.
- Try to answer every question.
- Check your answers if you have time at the end.
- Good luck with your examination.

Turn over ▶

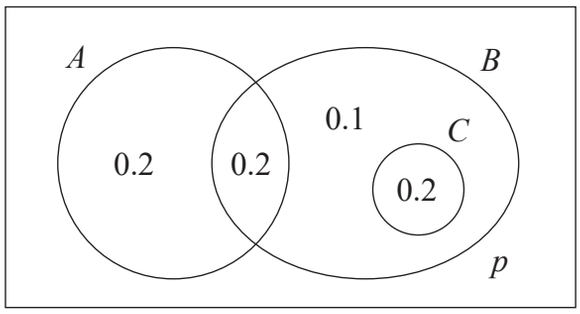
P66806A

1/1/1/1/1/

P 6 6 8 0 6 A 0 1 1 6

Pearson

1.



The Venn diagram, where p is a probability, shows the 3 events A, B and C with their associated probabilities.

(a) Find the value of p.

(1)

(b) Write down a pair of mutually exclusive events from A, B and C.

(1)

a) all probabilities add to 1

$0.2 + 0.2 + 0.1 + 0.2 + p = 1$

$0.7 + p = 1$

$p = 0.3$

b) Mutually exclusive means that A, B and C have nothing in common i.e. no overlap

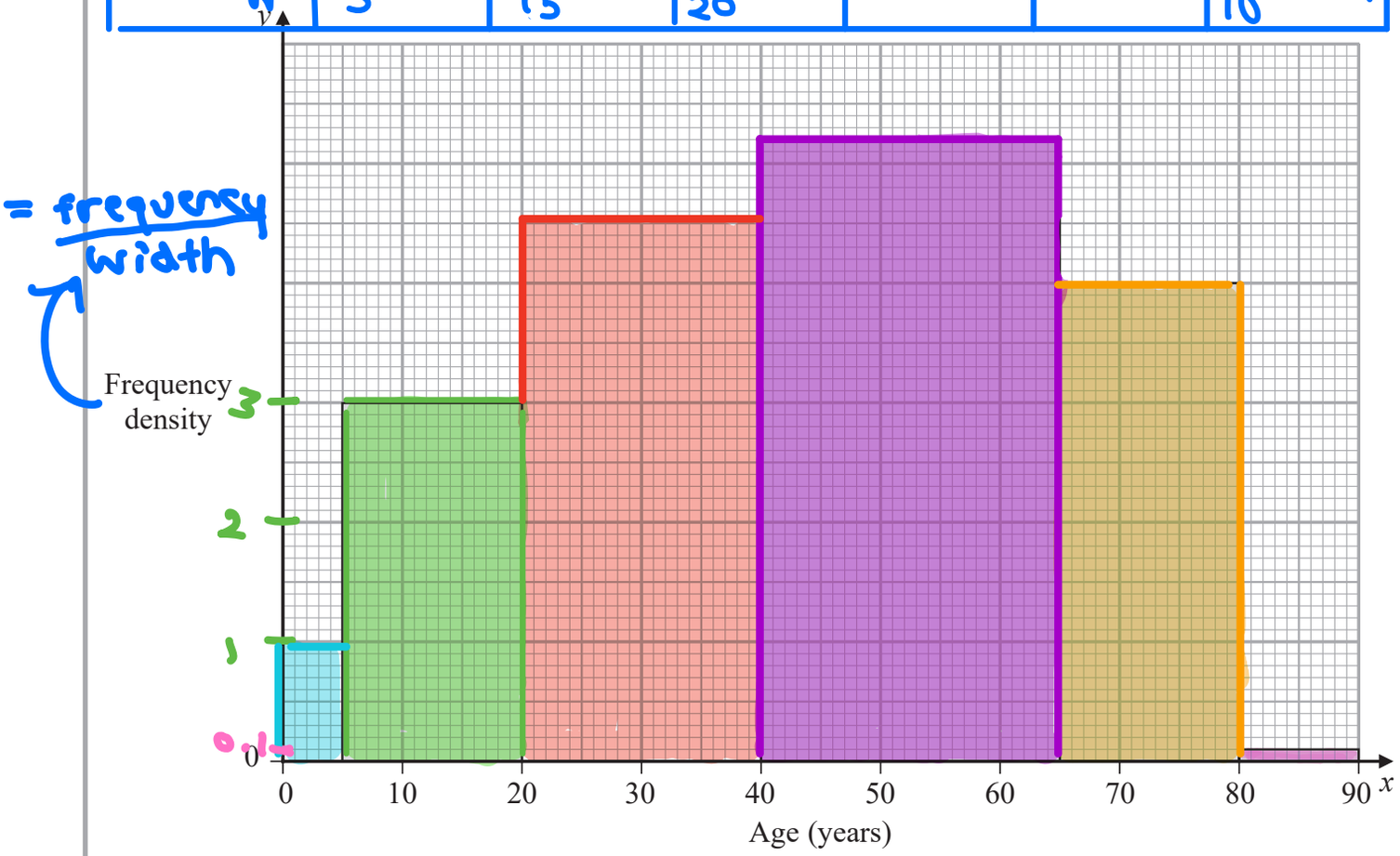
$P(A \cap B \cap C) = 0$

A and C are mutually exclusive

2. The partially completed table and partially completed histogram give information about the ages of passengers on an airline.

There were no passengers aged 90 or over.

Width	5 − 0 = 5	20 − 5 = 15	40 − 20 = 20	65 − 40 = 25	80 − 65 = 15	90 − 80 = 10
Age (x years)	$0 \leqslant x < 5$	$5 \leqslant x < 20$	$20 \leqslant x < 40$	$40 \leqslant x < 65$	$65 \leqslant x < 80$	$80 \leqslant x < 90$
Frequency	5	45	90	130	60	1
FD = F/W	5/5 = 1	45/15 = 3	90/20 = 4.5			1/10 = 0.1



= frequency / width

(a) Complete the histogram.

(3)

(b) Use linear interpolation to estimate the median age.

(4)

An outlier is defined as a value greater than $Q_3 + 1.5 \times$ interquartile range.

Given that $Q_1 = 27.3$ and $Q_3 = 58.9$

(c) determine, giving a reason, whether or not the oldest passenger could be considered as an outlier.

(2)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 2 continued

a) use the pink or green bars to get the scale since we are given the frequency and they're drawn for us

Way 1:

$FD = \frac{E}{W} = \frac{45}{15} = 3$

So each - on y axis represents $\frac{3}{3} = 1$

Way 2:

$FD = \frac{E}{W} = \frac{1}{10} = 0.1$

So each - on y axis represents 0.1

Let's find the missing frequencies in the table. The frequency is the area of the rectangles

40-65: Area = 25(5.2) = 130

65-80: Area = 15(4) = 60

b) Let's use the table

Age	Frequency
$0 \le w < 5$	5
$5 \le w < 20$	45
$20 \le w < 40$	90
$40 \le w < 65$	130
$65 \le w < 80$	60
$80 \le w < 90$	1

Way 1: Shorter Method

$$median = \frac{n}{2} = \frac{331}{2} = 165.5^{th} \text{ value}$$

See where 165.5 would insert in the cf column and drop down to next row

Age	Frequency	UCB	cf
$0 \le w < 5$	5	5	5
$5 \le w < 20$	45	20	50
$20 \le w < 40$	90	40	140
$40 \le w < 65$	130	65	270
$65 \le w < 80$	60	80	330
$80 \le w < 90$	1	90	331

165.5

Apply the formula: $\text{LCB} + \frac{\text{how many in}}{\text{group total}} \times \text{class width}$

$$40 + \frac{165.5 - 140}{130} \times (65 - 40) = 44.9039$$

$$= 44.9$$

Way 2: Longer Method

$$median = \frac{n}{2} = \frac{331}{2} = 165.5^{th} \text{ value}$$

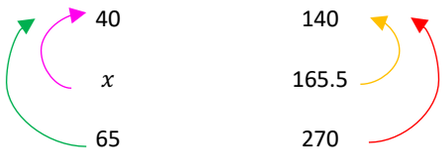
See where 140.5 would insert in the cf column

Find where the corresponding x value would be in the UCB column and call it x

Upper Class Boundary	cf
5	5
20	50
40	140
65	270
80	330
90	331

x 165.5

zoom in on the rows above and below of where we insert



We subtract the distances indicated above

$$\frac{x - 40}{65 - 40} = \frac{165.5 - 140}{270 - 140}$$

$$\frac{x - 40}{25} = \frac{25.5}{130}$$

$$x - 40 = 25\left(\frac{25.5}{130}\right)$$

$$x - 40 = 4.9038$$

$$x = 44.9038$$

$$x = 44.9$$

Median = 44.9

c) outlier > UQ + 1.5 IQR

> 58.9 + 1.5 (58.9 - 27.3)

> 58.9 + 47.4

> 106.3

so any value greater than 106.3 is an outlier

90 is the largest possible value and 90 < 106.3 so oldest passenger is not an outlier

3. Helen is studying one of the qualitative variables from the large data set for Heathrow from 2015.

She started with the data from 3rd May and then took every 10th reading.

There were only 3 different outcomes with the following frequencies

Outcome	*A*	*B*	*C*
Frequency	16	2	1

(a) State the sampling technique Helen used.

(1)

(b) From your knowledge of the large data set

(i) suggest which variable was being studied,

(ii) state the name of outcome *A*.

(2)

George is also studying the same variable from the large data set for Heathrow from 2015. He started with the data from 5th May and then took every 10th reading and obtained the following

Outcome	*A*	*B*	*C*
Frequency	16	1	1

Helen and George decided they should examine all of the data for this variable for Heathrow from 2015 and obtained the following

Outcome	*A*	*B*	*C*
Frequency	155	26	3

(c) State what inference Helen and George could reliably make from their original samples about the outcomes of this variable at Heathrow, for the period covered by the large data set in 2015.

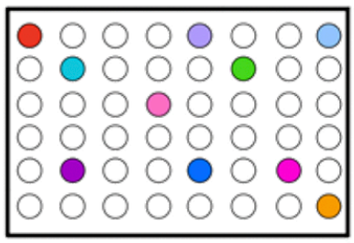
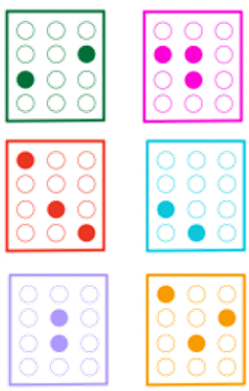
(1)

Q Let's look at the difference between simple random sample, stratified and systematic sampling

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Simple Random Sample	Stratified	Systematic
Choose a subset of participants from the population. 	Split the population into smaller groups called strata (subgroups) based on common characteristics/shared attributes and then take a simple random sample from each strata.  This ensures that every characteristic is properly represented in the sample	Choose subjects in a systematic, orderly/logical way by sampling every k^{th} element/value 

so in this case we have Systematic sampling

b) i) Daily mean wind speed

ii) Light

c) Variable A occurs most (around 80% - 90% of the time)

(Total for Question 3 is 4 marks)

4. A nursery has a sack containing a large number of coloured beads of which 14% are coloured red.

Aliya takes a random sample of 18 beads from the sack to make a bracelet.

(a) State a suitable binomial distribution to model the number of red beads in Aliya's bracelet.

(1)

(b) Use this binomial distribution to find the probability that

(i) Aliya has just 1 red bead in her bracelet,

(ii) there are at least 4 red beads in Aliya's bracelet.

(3)

(c) Comment on the suitability of a binomial distribution to model this situation.

(1)

After several children have used beads from the sack, the nursery teacher decides to test whether or not the proportion of red beads in the sack has changed.
She takes a random sample of 75 beads and finds 4 red beads.

(d) Stating your hypotheses clearly, use a 5% significance level to carry out a suitable test for the teacher.

c.v. if an so can't be one tailed

(4)

(e) Find the *p*-value in this case.

(1)

a)

let X be the number of red beads in Aliya's bracelet

$X \sim B(18, 0.14)$

↑ number of trials

↳ probability

b)

i) $P(X = 1)$

Use binompd on calculator with X value 1

0.194

ii) $P(X \geq 4)$

Way 1: use Classwiz calculator

This needs to be turned into a $\leq$

$P(X \geq 4)$ means $X = 4, 5, 6, 7, 8, 9, \ldots 18$

All probabilities add to 1

so $P(X \geq 4)$ is the same as saying

$= 1 - P(X \leq 3)$

All options

We subtract what we don't want from 1 in order to get what we want

Use binomcd on calculator to get $P(X \leq 3)$. x value is 3

$= 1 - 0.76184$

$= 0.238$

Way 2: Using Casio FX-CG50 calc.

$P(X \geq 4)$

We can use the calculator straight away

Use binomcd Lower: 4 Upper: 18 We get 0.238

c) The binomial distribution requires the probability to be 0.14 to be constant (not change). This means that a large number of beads in the sack is needed to ensure that removing 18 beads does not appreciably affect this probability. If this is the case, then the model could be suitable

d)

let x be the number of red beads in the sample

$X \sim B(75, 0.14)$

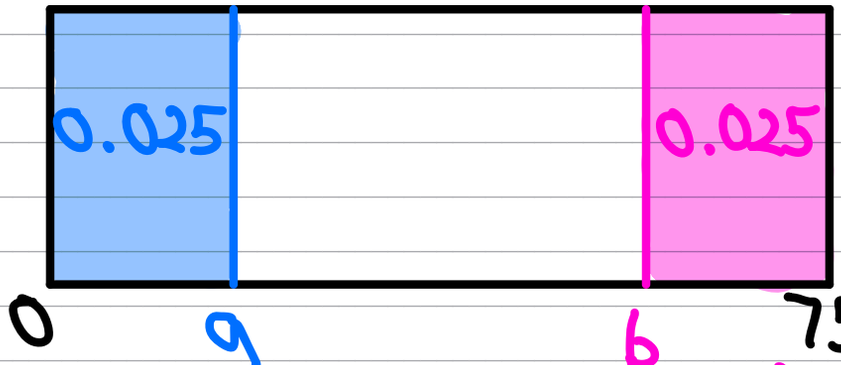
$H_0: p = 0.14$

$H_1: p \neq 0.14$

Way 1: critical value method

We reject when in either of the shaded regions (i.e in the critical region).

Question 4 continued



To find a:
$P(X \leq a) \leq 0.025$
use tables or calc
If calc use
binomcd with
List or invbinom

To find b:
$P(X \geq b) \leq 0.025$
$1 - P(X \leq b-1) \leq 0.025$
$P(X \leq b-1) \geq 0.975$
use tables or calc
If calc use
binomcd with
List or invbinom

≤	
Casio Classwiz	**Casio FX-CG50**
Menu 7:Distribution BinomialCD List Write all X's under first column (write all the way to n or if n is big guess high enough) find n value corresponding to being closest to $\leq \alpha$	2:Statistics F5: Dist F5:Binomial F3 InvB Data: variable Area:α Numtrial: n P: probability

Calculator gives
$a = 4$

≥	
Casio Classwiz	**Casio FX-CG50**
Turn into ≤ $1-P(X \leq CV-1) \leq \alpha$ Re-arrange $P(X \leq CV-1) \geq 1-\alpha$ Menu 7:Distribution BinomialCD List Write all n's under first column (guess high enough) Find n value that corresponds to being closest to $\geq 1-\alpha$ This and is CV−1 Add 1 to get CV	Turn into ≤ $1-P(X \leq CV-1) \leq \alpha$ Rearrange $P(X \leq CV-1) \geq 1-\alpha$ 2:Statistics F5: Dist F5:Binomial F3 InvB Data: variable Area: $1-\alpha$ Numtrial: n P: probability This and is CV−1. Add 1 to get CV

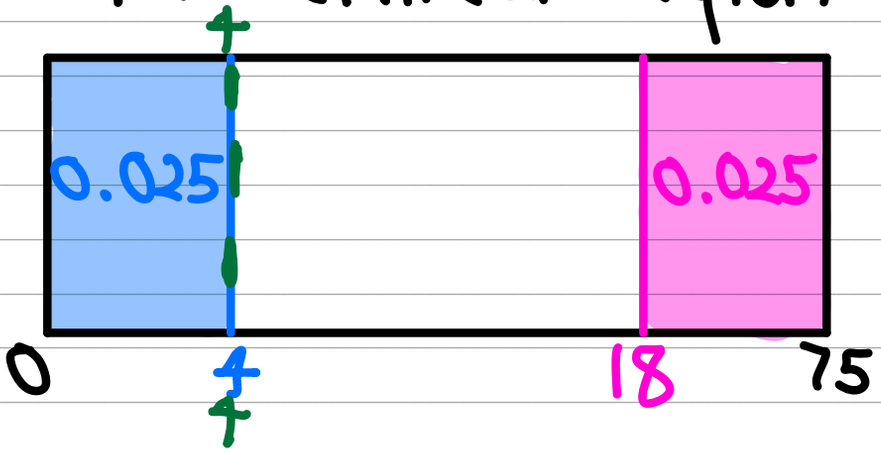
Calculator gives
$b-1 = 17 \Rightarrow b = 18$

So critical regions are

- $X \leq 4$ i.e $0 \leq X \leq 4$
- $X \geq 18$ i.e $18 \leq X \leq 75$

(Total for Question 4 is 10 marks)

Now we need to check whether t is in the critical region



t is in the critical region (right on the boundary)

So we reject H_0

∴ there is sufficient evidence to reject the null hypothesis and we can conclude that the proportion of red beads has changed

Note: The conclusion always talks about whether accept or reject H_0 and then we specify what that means in context after

Way 2: P Value method

Find $P(X \leq 4)$

There are 2 ways to do this depending on which calculator you have.

(Total for Question 4 is 10 marks)

Way 1: Use classwiz calculator

Use binomcd
X value = 4

= 0.01506

Way 2: Using Casio FX-CG50 calc.

P(X ≤ 4)

We can use the calculator straight away

Use binomcd
Lower = 0
Upper = 4

= 0.01506

0.01506 < 0.025 so reject Ho

↑ half of significance level since 2 tailed

OR: 2(0.01506) = 0.03012 < 0.05 so reject Ho

(same conclusion as in Way 1)

c) 0.03 found above

5. Two bags, **A** and **B**, each contain balls which are either red or yellow or green.

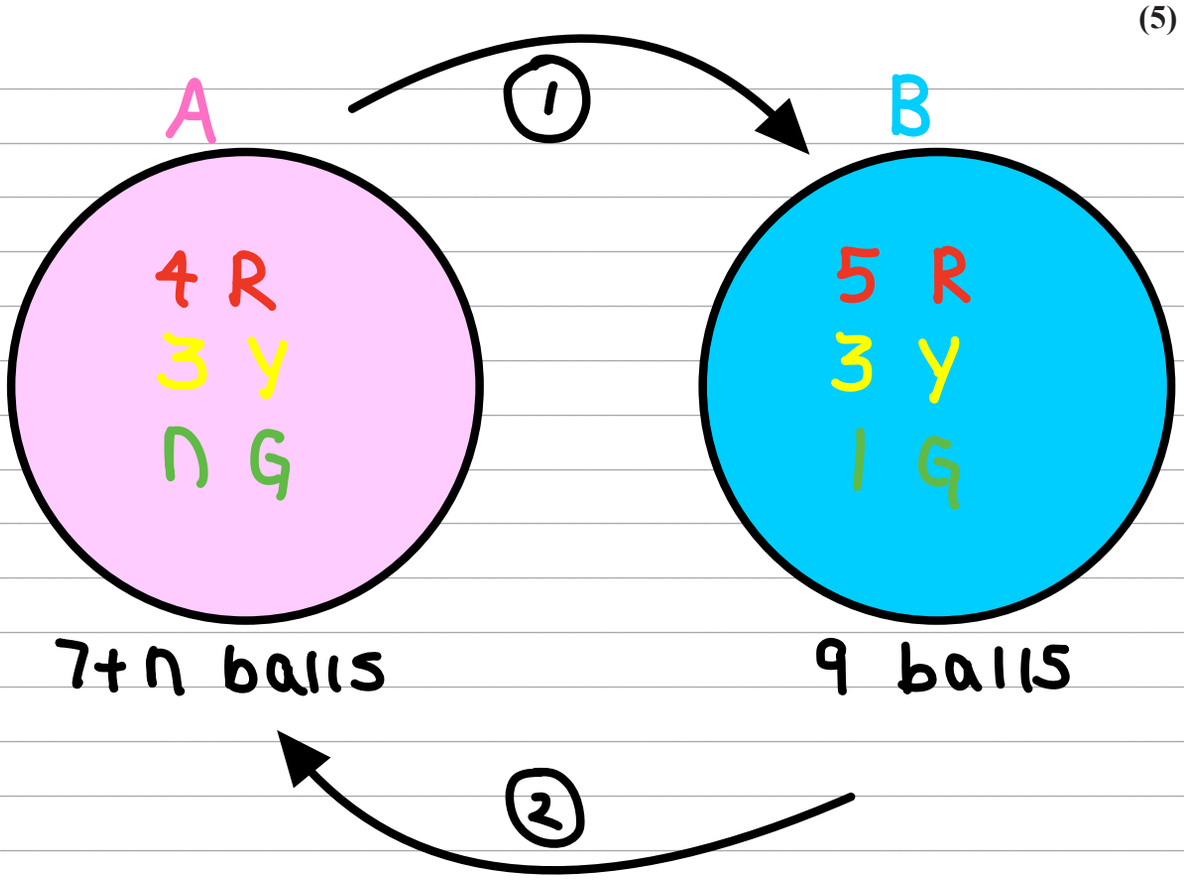
Bag **A** contains 4 red, 3 yellow and n green balls.
Bag **B** contains 5 red, 3 yellow and 1 green ball.

A ball is selected at random from bag **A** and placed into bag **B**.
A ball is then selected at random from bag **B** and placed into bag **A**.

The probability that bag **A** now contains an equal number of red, yellow and green balls is p.

Given that $p > 0$, find the possible values of n and p.

(5)



We need to end up with the same number of each colour in bag A (the pink bag).

The only way to do this is to get 3 of each colour in bag A or 4 of each colour in bag A

The only options for n to achieve this is for n to be 2 or 5.

Start:

Bag A	Bag B
4	5
3	3
n	1

Case 1: $n = 2$
Take one red ball from bag A and put it into B and then take one green ball from bag B and put it into bag A

Pick from bag A	
Bag A	Bag B
4	5
3	3
2	1

$\frac{4}{9}$

Pick from bag B	
Bag A	Bag B
3	6
3	3
2	1

$\frac{1}{10}$

Result	
Bag A	Bag B
3	6
3	3
3	0

$$\frac{4}{9} \times \frac{1}{10} = \frac{2}{45}$$

Case 2: $n = 5$
Take one green ball from bag A and put it into B and then take one yellow ball from bag B and put it into bag A

Pick from bag A	
Bag A	Bag B
4	5
3	3
5	1

$\frac{5}{12}$

Pick from bag B	
Bag A	Bag B
4	5
3	3
4	2

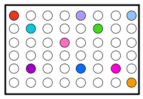
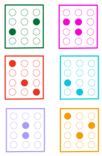
$\frac{3}{10}$

Result	
Bag A	Bag B
4	5
4	2
4	2

$$\frac{5}{12} \times \frac{3}{10} = \frac{1}{8}$$

So $n = 2, p = \frac{2}{25}$ or $n = 5, p = \frac{1}{8}$

≥	
Casio Classwiz	**Casio FX-CG50**
Turn into ≤ $1-\mathrm{P}(X \leq CV-1) \leq \alpha$ Re-arrange $P(X \leq CV-1) \geq 1-\alpha$ Menu 7:Distribution BinomialCD List Write all n's under first column (guess high enough) Find n value that corresponds to being closest to $\geq 1-\alpha$ This and is CV−1 Add 1 to get CV	Turn into ≤ $1-\mathrm{P}(X \leq CV-1) \leq \alpha$ Rearrange $P(X \leq CV-1) \geq 1-\alpha$ 2:Statistics F5: Dist F5:Binomial F3 InvB Data: variable Area: $1-\alpha$ Numtrial: n P: probability This and is CV−1. Add 1 to get CV

Simple Random Sample	Stratified	Systematic
Choose a subset of participants from the population. 	Split the population into smaller groups called strata (subgroups) based on common characteristics/shared attributes and then take a simple random sample from each strata.  This ensures that every characteristic is properly represented in the sample	Choose subjects in a systematic, orderly/logical way by sampling every k^{th} element/value