Dr Oliver Mathematics
Further Mathematics
Populations and Samples
Past Examination Questions

This booklet consists of 19 questions across a variety of examination topics.
The total number of marks available is 109.

1. A magazine has a large number of subscribers who each pay a membership fee that is due on January 1st each year. Not all subscribers pay their fee by the due date. Based on correspondence from the subscribers, the editor of the magazine believes that 40% of subscribers wish to change the name of the magazine. Before making this change the editor decides to carry out a sample survey to obtain the opinions of the subscribers. He uses only those members who have paid their fee on time.

   (a) Define the population associated with the magazine. (1)

   Solution
   All of the subscribers to the magazine.

   (b) Suggest a suitable sampling frame for the survey. (1)

   Solution
   A list of all members that had paid their subscriptions.

   (c) Identify the sampling units. (1)

   Solution
   The members who have paid.

   (d) Give one advantage and one disadvantage that would have resulted from the editor using a census rather than a sample survey. (2)

   Solution
   Advantage: every single member of the population is used, it is unbiased, or total accuracy.
   Disadvantage: time consuming to obtain the data and then to analyse it, it is costly, and it is often difficult to ensure that the whole population is surveyed.

2. Explain what you understand by

(a) a sampling frame, (1)

> **Solution**
> When the sampling units within a population are individually named or numbered to form a list then this list of sampling units is called a sampling frame.

(b) a statistic. (2)

> **Solution**
> A statistic is a random variable calculated solely from the observation in a random sample and it does not involve any unknown parameters.

3. (a) Explain what you understand by (i) a population and (ii) a sampling frame. (2)

> **Solution**
> (i) A population is a collection of individual persons or items.
> (ii) When the sampling units within a population are individually named or numbered to form a list then this list of sampling units is called a sampling frame.

The population and the sampling frame may not be the same.

(b) Explain why this might be the case. (1)

> **Solution**
> It is not always possible to keep the lists up-to-date.

(c) Give an example, justifying your choices, to illustrate when you might use

(i) a census, (2)

> **Solution**
> e.g., a school: the population is known and is easily accessed (small but easily listed population).

(ii) a sample. (2)

> **Solution**
> e.g., a university: the population is known but too time consuming to interview all of them (large but not easily listed population).

4. Explain what you understand by

(a) a sampling unit, (1)

Solution
The individual units of a population are known as sampling units.

(b) a sampling frame, (1)

Solution
When the sampling units within a population are individually named or numbered to form a list then this list of sampling units is called a sampling frame.

(c) a sampling distribution. (2)

Solution
All possible samples are chosen from a population; the values of a statistic and the associated probabilities is a sampling distribution.

5. A bag contains a large number of coins. Half of them are 1p coins, one third are 2p coins, and the remainder are 5p coins.

(a) Find the mean and variance of the value of the coins. (4)

Solution
Now,

$$1 - \frac{1}{2} - \frac{1}{3} = \frac{1}{6}$$

are 5p coins.

$$\begin{aligned}\text{Mean} &= \frac{1}{2} \times 1 + \frac{1}{3} \times 2 + \frac{1}{6} \times 5 \\ &= \frac{1}{2} + \frac{2}{3} + \frac{5}{6} \\ &= \underline{\underline{2}}\end{aligned}$$

and

$$\begin{aligned}\text{variance} &= \frac{1}{2} \times 1^2 + \frac{1}{3} \times 2^2 + \frac{1}{6} \times 5^2 - 2^2 \\ &= \frac{1}{2} + \frac{4}{3} + \frac{25}{6} - 4 \\ &= \underline{\underline{2}}.\end{aligned}$$

A random sample of 2 coins is chosen from the bag.

(b) List all the possible samples that can be drawn. (3)

Solution

$(1,1), (1,2), (1,5), (2,1), (2,2), (2,5), (5,1), (5,2), (5,5)$

(c) Find the sampling distribution of the mean value of these samples. (6)

Solution

$\text{P}(\overline{X} = 1) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$
$\text{P}(\overline{X} = \frac{3}{2}) = 2 \times \frac{1}{2} \times \frac{1}{3} = \frac{1}{3}$
$\text{P}(\overline{X} = 2) = \frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$
$\text{P}(\overline{X} = 3) = 2 \times \frac{1}{2} \times \frac{1}{6} = \frac{1}{6}$
$\text{P}(\overline{X} = \frac{7}{2}) = 2 \times \frac{1}{3} \times \frac{1}{6} = \frac{1}{9}$
$\text{P}(\overline{X} = 5) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$

$\overline{x}$	1	$\frac{3}{2}$	2	3	$\frac{7}{2}$	5
$\text{P}(\overline{X} = \overline{x})$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{9}$	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{1}{36}$

6. Before introducing a new rule the secretary of a golf club decided to find out how members might react to this rule.

(a) Explain why the secretary decided to take a random sample of club members rather than ask all the members. (1)

Solution

It is easier, cheaper, or takes less time than taking census.

(b) Suggest a suitable sampling frame. (1)

Solution

The full membership list are sampling frame.

(c) Identify the sampling units. (1)

Solution

The club members list are sampling units.

7. (a) Define a statistic. (2)

Solution
A statistic is a random variable calculated solely from the observation in a random sample and it does not involve any unknown parameters.

A random sample X_1, X_2, ..., X_n is taken from a population with unknown mean μ.

(b) For each of the following state whether or not it is a statistic.

(i) $\dfrac{X_1 + X_4}{2}$, (1)

Solution
Yes.

(ii) $\dfrac{\Sigma X^2}{n} - \mu^2$. (1)

Solution
No.

8. A bag contains a large number of coins: 75% are 10p coins and 25% are 5p coins.
A random sample of 3 coins is drawn from the bag.
Find the sampling distribution for the median of the values of the 3 selected coins.

Solution

$$\begin{aligned}
\text{P(median is 5)} &= \binom{3}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right) + \left(\frac{1}{4}\right)^3 \\
&= \frac{9}{64} + \frac{1}{64} \\
&= \frac{10}{64} \\
&= \frac{5}{32}.
\end{aligned}$$

$$\begin{aligned}
\text{P(median is 10)} &= \binom{3}{2}\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^2 + \left(\frac{3}{4}\right)^3 \\
&= \frac{27}{64} + \frac{27}{64} \\
&= \frac{54}{64} \\
&= \frac{27}{32}.
\end{aligned}$$

$\overline{x}$	5	10
$\mathrm{P}(\overline{X} = \overline{x})$	$\frac{5}{32}$	$\frac{27}{32}$

9. (a) Explain what you understand by a census. (1)

Solution
The investigation is a census if the information is obtained from all members of the population.

Each cooker produced at GT Engineering is stamped with a unique serial number. GT Engineering produces cookers in batches of 2000. Before selling them, they test a random sample of 5 to see what electric current overload they will take before breaking down.

(b) Give one reason, other than to save time and cost, why a sample is taken rather than a census. (1)

Solution
There will be no cookers left if they take a census.

(c) Suggest a suitable sampling frame from which to obtain this sample. (1)

Solution
The list of the unique serial numbers of the cookers.

(d) Identify the sampling units. (1)

Solution
A cooker.

10. A random sample X_1, X_2, ..., X_n is taken from a population with unknown mean μ and unknown variance σ^2. A statistic Y is based on this sample.

(a) Explain what you understand by the statistic Y. (2)

Solution
A statistic is a random variable calculated solely from the observation in a random sample and it does not involve any unknown parameters.

(b) Explain what you understand by the sampling distribution of Y. (1)

Solution

If we repeatedly take samples from a population and calculate the same statistic each time there is a range of values that the statistic can take. The statistic will have its own distribution which we call the sampling distribution.

(c) State, giving a reason which of the following is not a statistic based on this sample. (2)

(i) $\displaystyle\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n}$

(ii) $\displaystyle\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2$

(iii) $\displaystyle\sum_{i=1}^{n} X_i^2$

Solution

(i) Yes: a statistic
(ii) No: as it contains μ and σ
(iii) Yes: a statistic.

11. A bag contains a large number of coins. It contains only 1p and 2 p coins in the ratio 1 : 3.

(a) Find the mean μ and the variance σ^2 of the values of this population of coins. (3)

Solution

$$\begin{aligned}\mu &= \frac{1}{4} \times 1 + \frac{3}{4} \times 2 \\ &= \frac{1}{4} + \frac{3}{2} \\ &= \underline{\underline{\frac{7}{4}}}\end{aligned}$$

and

$$\begin{aligned}\sigma^2 &= \frac{1}{4} \times 1^2 + \frac{3}{4} \times 2^2 - \left(\frac{7}{4}\right)^2 \\ &= \frac{13}{4} - \frac{49}{16} \\ &= \underline{\underline{\frac{3}{16}}}.\end{aligned}$$

A random sample of size 3 is taken from the bag.

(b) List all the possible samples. (2)

Solution

(1, 1, 1), (1, 1, 2), (1, 2, 1), (2, 1, 1), (1, 2, 2), (2, 1, 2), (2, 2, 1), (2, 2, 2)

(c) Find the sampling distribution of the mean value of the samples. (6)

Solution

$P(\overline{X} = 1) = \left(\frac{1}{4}\right)^3 = \frac{1}{64}$
$P(\overline{X} = \frac{4}{3}) = 3\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right) = \frac{9}{64}$
$P(\overline{X} = \frac{5}{3}) = 3\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^2 = \frac{27}{64}$
$P(\overline{X} = 2) = \left(\frac{3}{4}\right)^3 = \frac{27}{64}$

$\overline{x}$	1	$\frac{4}{3}$	$\frac{5}{3}$	2
$P(\overline{X} = \overline{x})$	$\frac{1}{64}$	$\frac{9}{64}$	$\frac{27}{64}$	$\frac{27}{64}$

12. Explain what you understand by

(a) a population, (1)

Solution

A population is a collection of individual persons or items.

(b) a statistic. (1)

Solution

A statistic is a random variable calculated solely from the observation in a random sample and it does not involve any unknown parameters.

A researcher took a sample of 100 voters from a certain town and asked them who they would vote for in an election. The proportion who said they would vote for Dr Smith was 35%.

(c) State the population and the statistic in this case. (2)

Solution

Population: the voters in the town.
Statistic: those who voted for Dr Smith.

(d) Explain what you understand by the sampling distribution of this statistic. (1)

> **Solution**
> If we repeatedly take samples from a population and calculate the same statistic each time there is a range of values that the statistic can take. The statistic will have its own distribution which we call the sampling distribution.

13. A factory produces components. Each component has a unique identity number and it is assumed that 2% of the components are faulty. On a particular day, a quality control manager wishes to take a random sample of 50 components.

(a) Identify a sampling frame. (1)

> **Solution**
> The list of the unique identity number of the components.

The statistic F represents the number of faulty components in the random sample of size 50.

(b) Specify the sampling distribution of F. (2)

> **Solution**
> Let F represent the number of unique identity number $\therefore \underline{\underline{F \sim \mathrm{B}(50, 0.02)}}$.

14. A bag contains a large number of balls: 65% are numbered 1 and 35% are numbered 2. A random sample of 3 balls is taken from the bag. Find the sampling distribution for the range of the numbers on the 3 selected balls. (6)

> **Solution**
> There are $(1, 1, 1)$, $(1, 1, 2)$, $(1, 2, 1)$, $(2, 1, 1)$, $(1, 2, 2)$, $(2, 1, 2)$, $(2, 2, 1)$, and $(2, 2, 2)$ to choose from.
>
> $$\begin{aligned}\mathrm{P}(\text{range is } 0) &= \left(\frac{65}{100}\right)^3 + \left(\frac{35}{100}\right)^3 \\ &= \frac{274\,625}{1\,000\,000} + \frac{42\,875}{1\,000\,000} \\ &= \frac{317\,500}{1\,000\,000} \\ &= \frac{127}{400}.\end{aligned}$$

$$
\begin{aligned}
\mathrm{P}(\text{range is } 1) &= \binom{3}{2}\left(\frac{65}{100}\right)\left(\frac{35}{100}\right)^2 + \binom{3}{2}\left(\frac{65}{100}\right)^2\left(\frac{35}{100}\right) \\
&= \frac{238\,875}{1\,000\,000} + \frac{443\,625}{1\,000\,000} \\
&= \frac{682\,500}{1\,000\,000} \\
&= \frac{273}{400}.
\end{aligned}
$$

$\overline{x}$	0	1
$\mathrm{P}(\overline{X} = \overline{x})$	$\frac{127}{400}$	$\frac{273}{400}$

15. A bag contains a large number of 1p, 2p, and 5p coins: 50% are 1p coins, 20% are 2p coins, and 30% are 5p coins A random sample of 3 coins is chosen from the bag.

(a) List all the possible samples of size 3 with median 5p. (2)

Solution

$(1, 5, 5)$, $(5, 1, 5)$, $(5, 5, 1)$, $(2, 5, 5)$, $(5, 2, 5)$, $(5, 5, 2)$, $(5, 5, 5)$.

(b) Find the probability that the median value of the sample is 5p. (4)

Solution

$$
\begin{aligned}
\mathrm{P}(\text{median 5p}) &= \binom{3}{2}(0.3)^2(0.5) + \binom{3}{2}(0.3)^2(0.2) + (0.3)^3 \\
&= 0.135 + 0.054 + 0.027 \\
&= \underline{\underline{0.216}}.
\end{aligned}
$$

(c) Find the sampling distribution of the median of samples of size 3. (5)

Solution

$$
\begin{aligned}
\mathrm{P}(\text{median 1p}) &= \binom{3}{2}(0.5)^2(0.2) + \binom{3}{2}(0.5)^2(0.3) + (0.5)^3 \\
&= 0.15 + 0.225 + 0.125 \\
&= 0.5
\end{aligned}
$$

and

$$\begin{aligned}\text{P(median 2p)} &= \binom{3}{2}(0.2)^2(0.5) + \binom{3}{2}(0.2)^2(0.3) + (0.2)^3 + 6(0.5)(0.2)(0.3)\\ &= 0.06 + 0.036 + 0.008 + 0.18\\ &= 0.284.\end{aligned}$$

$\overline{x}$	1	2	5
$\text{P}(\overline{X} = \overline{x})$	0.5	0.284	0.216

16. A bag contains a large number of counters. A third of the counters have a number 5 on them and the remainder have a number 1.

A random sample of 3 counters is selected.

(a) List all possible samples. (2)

Solution
$(1,1,1), (1,1,5), (1,5,1), (5,1,1), (1,5,5), (5,1,5), (1,5,5), (5,5,5)$.

(b) Find the sampling distribution for the range. (3)

Solution
$\text{P}(\overline{X} = 0) = \left(\frac{2}{3}\right)^3 + \left(\frac{1}{3}\right)^3 = \frac{1}{3}$
$\text{P}(\overline{X} = 4) = 3\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right) + 3\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^2 = \frac{2}{3}$

$\overline{x}$	0	4
$\text{P}(\overline{X} = \overline{x})$	$\frac{1}{3}$	$\frac{2}{3}$

17. A bag contains a large number of counters. Each counter has a single digit number on it and the mean of all the numbers in the bag is the unknown parameter μ. The number 2 is on 40% of the counters and the number 5 is on 25% of the counters. All the remaining counters have numbers greater than 5 on them. A random sample of 10 counters is taken from the bag.

State whether or not each of the following is a statistic.

(a) S = the sum of the numbers on the counters in the sample, (1)

Solution
Yes.

(b) D = the difference between the highest number in the sample and μ, (1)

Solution
No.

(c) F = the number of counters in the sample with a number 5 on them. (1)

Solution
Yes.

18. A bag contains a large number of 10p, 20p, and 50p coins in the ratio 1 : 2 : 2. (7)

A random sample of 3 coins is taken from the bag.

Find the sampling distribution of the median of these samples.

Solution
$\text{P(median 10p)} = (0.2)^3 + \binom{3}{2}(0.2)^2(0.4) + \binom{3}{2}(0.2)^2(0.4) = 0.104$
$\text{P(median 20p)} = \binom{3}{2}(0.4)^2(0.2) + (0.4)^3 + \binom{3}{2}(0.4)^2(0.4) + 6(0.2)(0.4)(0.4) = 0.544$
$\text{P(median 50p)} = \binom{3}{2}(0.4)^2(0.2) + \binom{3}{2}(0.4)^2(0.4) + (0.4)^3 = 0.352$

$\overline{x}$	10	20	50
$\text{P}(\overline{X} = \overline{x})$	0.104	0.544	0.352

19. A bag contains a large number of counters with one of the numbers 4, 6 or 8 written on each of them in the ratio 5 : 3 : 2 respectively.

A random sample of 2 counters is taken from the bag.

(a) List all the possible samples of size 2 that can be taken. (2)

Solution
$(4,4)$, $(4,6)$, $(4,8)$, $(6,4)$, $(6,6)$, $(6,8)$, $(8,4)$, $(8,6)$, $(8,8)$.

The random variable M represents the mean value of the 2 counters. Given that

$$\mathrm{P}(M = 4) = \tfrac{1}{4} \text{ and } \mathrm{P}(M = 8) = \tfrac{1}{25},$$

(b) find the sampling distribution for M. (5)

Solution

$\mathrm{P}(4, 6 \text{ or } 6, 4) = 2 \times \frac{1}{2} \times \frac{3}{10} = \frac{3}{10}$
$\mathrm{P}(4, 8 \text{ or } 8, 4 \text{ or } 6, 6) = 2 \times \frac{1}{2} \times \frac{1}{5} + \left(\frac{3}{10}\right)^2 = \frac{29}{100}$
$\mathrm{P}(8, 6 \text{ or } 6, 8) = 2 \times \frac{3}{10} \times \frac{1}{5} = \frac{3}{25}$

$\overline{x}$	4	5	6	7	8
$\mathrm{P}(\overline{X} = \overline{x})$	$\frac{1}{4}$	$\frac{3}{10}$	$\frac{29}{100}$	$\frac{3}{25}$	$\frac{1}{25}$

A sample of n sets of 2 counters is taken. The random variable Y represents the number of these n sets that have a mean of 8.

(c) Calculate the minimum value of n such that $\mathrm{P}(Y \geqslant 1) > 0.9$. (3)

Solution

$$\begin{aligned}
\mathrm{P}(Y \geqslant 1) > 0.9 &\Rightarrow 1 - \mathrm{P}(Y = 0) > 0.9 \\
&\Rightarrow \mathrm{P}(Y = 0) < 0.1 \\
&\Rightarrow \left(\tfrac{24}{25}\right)^n < 0.1 \\
&\Rightarrow n \ln \tfrac{24}{25} < \ln 0.1 \\
&\Rightarrow n > \frac{\ln 0.1}{\ln \frac{24}{25}} \\
&\Rightarrow n > 56.405\ldots
\end{aligned}$$

so $\underline{\underline{n = 57}}$.