MS04 Support Material

Estimators

For use with AQA A-level Mathematics Specification (6360)

1 Terminology

1.1 Sample Statistic

A numerical measure, calculated from a sample, is called a **sample statistic**.

Thus the following are examples of sample statistics:

$$\bar{x} \qquad s^2 \qquad r \qquad a \qquad b \qquad \hat{p}$$

1.2 Population parameter

A numerical value that defines a population is called a **population parameter**.

Thus the following are examples of population parameters:

$$\mu \qquad \sigma^2 \qquad \rho \qquad \alpha \qquad \beta \qquad p$$

It is important to remember that these are **not** random variables.

1.3 Estimator and Estimate

The **random variable** of a sample statistic is called an **estimator** for a population parameter.

A **value** of a sample statistic is called an **estimate** of a population parameter.

Thus $\bar{X}$ and S^2 are estimators for μ and σ^2, respectively, whereas $\bar{x}$ and s^2 are estimates.

1.4 Sampling Distribution

The distribution of an estimator is called its **sampling distribution**.

Thus, if $X \sim \mathrm{N}(\mu, \sigma^2)$, then the distribution $\mathrm{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$ would be the sampling distribution of the estimator, $\bar{X}$.

2 Properties of Estimators

2.1 Unbiased

The sample statistic T is an **unbiased** estimator for the population parameter θ when:

$$\mathrm{E}(T) = \theta$$

2.2 Consistent

The sample statistic T is a **consistent** estimator for the population parameter θ when:

$$\mathrm{V}(T) \to 0 \qquad \text{as} \qquad n \to \infty$$

2.3 Relative Efficiency

If R and T are two estimators for the parameter θ, then the **relative efficiency** of R relative to (with respect to) T is given by:

$$\mathrm{RE}\ (R \text{ wrt } T) = \frac{\left[\mathrm{V}(R)\right]^{-1}}{\left[\mathrm{V}(T)\right]^{-1}}$$

3 Exemplar Examination Questions – Questions

3.1 Question 1

A random variable X is distributed with mean μ and standard deviation σ. Three independent observations, X_1, X_2 and X_3, of the random variable X are to be taken. The combined statistic:

$$aX_1 + bX_2 + cX_3\,,$$

where a, b and c are constants, is to be used as an estimator for μ.

(a) Obtain expressions for the mean and variance of this combined statistic.

(b) Explain why it is desirable that a, b and c are chosen so that $a + b + c = 1$.

(c) When $a = b = c = \frac{1}{3}$, the combined statistic is denoted by $\overline{X}$.
When $a = b = \frac{2}{3}$ and $c = -\frac{1}{3}$, the combined statistic is denoted by T.

Calculate the relative efficiency of $\overline{X}$ with respect to T.

3.2 Question 2

Steel rods in a large batch have lengths that are distributed with mean μ and variance σ^2. Two trainees are set the task of estimating μ. Trainee A intends to take a random sample of 20 rods and calculate the mean length, $\overline{X}_A$, of the rods in the sample.

(a) Write down expressions for the mean and variance of $\overline{X}_A$.

(b) Trainee B intends to take a random sample of 5 rods and calculate the mean length, $\overline{X}_B$, of the rods in the sample. Their instructor suggests that they could obtain a better estimate of μ by combining their individual estimates. The trainees consider two possible combinations, $\overline{X}_M$ and $\overline{X}_W$, of their estimators, where:

$$\overline{X}_M = \tfrac{1}{2}\overline{X}_A + \tfrac{1}{2}\overline{X}_B \quad \text{and} \quad \overline{X}_W = \tfrac{4}{5}\overline{X}_A + \tfrac{1}{5}\overline{X}_B$$

(i) Show that the estimator $\overline{X}_M$ has mean μ and variance $\dfrac{\sigma^2}{16}$.

(ii) State why $\overline{X}_M$ is an unbiased estimator for μ.

(iii) Show that $\overline{X}_W$ is an unbiased estimator for μ.

(iv) Calculate the relative efficiency of $\overline{X}_M$ with respect to $\overline{X}_W$, and give a reason why $\overline{X}_W$ should be preferred to $\overline{X}_M$ as an estimator for μ.

3.3 Question 3

A manufacturing process produces a large number of items of which a proportion p are defective. Alan takes a random sample of n items from the output and records X defective items, whilst Betty independently takes a random sample of $2n$ items from the output and records Y defective items.

(a) Write down the mean and variance of **each** of X and Y.

(b) Their supervisor considers two possible estimators for p:

$$\hat{p}_1 = \frac{1}{2}\left(\frac{X}{n} + \frac{Y}{2n}\right) \quad \text{and} \quad \hat{p}_2 = \frac{X+Y}{3n}$$

(i) Show that **each** of these estimators for p is unbiased.

(ii) Show that the variance of $\hat{p}_1$ is $\dfrac{3p(1-p)}{8n}$.

(iii) Calculate the relative efficiency of $\hat{p}_1$ relative to $\hat{p}_2$.

3.4 Question 4

The random variable X has a rectangular distribution on the interval $0 \le x \le 2\theta$.

(a) State, in terms of θ, the variance of X.

Independent observations, X_1, X_2, X_3, …, X_n, are taken of the random variable X. The mean of such a sample of observations is denoted by $\bar{X}$.

(b) Write down, in terms of θ and *n*, expressions for the mean and variance of $\bar{X}$.

(c) Explain why $\bar{X}$ is:

(i) an unbiased estimator for θ;

(ii) a consistent estimator for θ.

(d) For a random sample of size 5, the median, M, is an unbiased estimator for θ with variance $\dfrac{\theta^2}{7}$.

For such a sample, calculate the relative efficiency of $\bar{X}$ with respect to M, and give a reason why $\bar{X}$ should be preferred to M as an estimator for θ.

(e) A particular random sample of 5 observations of X gave the following values.

$$1.4 \qquad 2.5 \qquad 0.2 \qquad 0.4 \qquad 1.3$$

(i) Use this sample to write down an inequality for 2θ, the upper boundary of the interval for X.

(ii) Obtain the values of $\bar{x}$ and *m* for this sample.

(iii) Explain which, of $\bar{x}$ and m, is the more appropriate estimate of θ in this case.

4 Exemplar Examination Questions – Answers

4.1 Question 1

(a)

$$\mathrm{E}(aX_1 + bX_2 + cX_3) = a\mathrm{E}(X_1) + b\mathrm{E}(X_2) + c\mathrm{E}(X_3)$$

$$= a\mu + b\mu + c\mu$$

$$= \underline{(a + b + c)\mu}$$

$$\mathrm{V}(aX_1 + bX_2 + cX_3) = a^2\mathrm{V}(X_1) + b^2\mathrm{V}(X_2) + c^2\mathrm{V}(X_3)$$

$$= a^2\sigma^2 + b^2\sigma^2 + c^2\sigma^2$$

$$= \underline{(a^2 + b^2 + c^2)\sigma^2}$$

(b)

$$a + b + c = 1 \quad \Rightarrow \quad \mathrm{E}(aX_1 + bX_2 + cX_3) = \mu$$

Thus combined statistic is an unbiased estimator for μ

(c)

$$\mathrm{V}(\bar{X}) = \left(\frac{1}{9} + \frac{1}{9} + \frac{1}{9}\right)\sigma^2 = \underline{\frac{\sigma^2}{3}}$$

$$\mathrm{V}(T) = \left(\frac{4}{9} + \frac{4}{9} + \frac{(-1)^2}{9}\right)\sigma^2 = \underline{\sigma^2}$$

$$\text{RE } (\bar{X} \text{ wrt } T) = \frac{\left[\mathrm{V}(\bar{X})\right]^{-1}}{\left[\mathrm{V}(T)\right]^{-1}} = \frac{\left[\frac{\sigma^2}{3}\right]^{-1}}{\left[\sigma^2\right]^{-1}} = \underline{3}$$

4.2 Question 2

(a)

$$E\left(\bar{X}_A\right) = \underline{\mu} \qquad V\left(\bar{X}_A\right) = \underline{\frac{\sigma^2}{20}}$$

(b)
(i)

$$E\left(\bar{X}_M\right) = \frac{1}{2}E\left(\bar{X}_A\right) + \frac{1}{2}E\left(\bar{X}_B\right) = \frac{1}{2}\mu + \frac{1}{2}\mu = \underline{\mu}$$

$$V\left(\bar{X}_M\right) = \frac{1}{4}V\left(\bar{X}_A\right) + \frac{1}{4}V\left(\bar{X}_B\right) = \frac{1}{4}\times\frac{\sigma^2}{20} + \frac{1}{4}\times\frac{\sigma^2}{5} = \underline{\frac{\sigma^2}{16}}$$

(ii)

$E\left(\bar{X}_M\right) = \mu \quad \Rightarrow \quad \bar{X}_M$ is unbiased

(iii)

$E\left(\bar{X}_W\right) = \frac{4}{5}E\left(\bar{X}_A\right) + \frac{1}{5}E\left(\bar{X}_B\right) = \frac{4}{5}\mu + \frac{1}{5}\mu = \underline{\mu} \quad \Rightarrow \quad \bar{X}_W$ is unbiased

$$V\left(\bar{X}_W\right) = \frac{16}{25}V\left(\bar{X}_A\right) + \frac{1}{25}V\left(\bar{X}_B\right) = \frac{16}{25}\times\frac{\sigma^2}{20} + \frac{1}{25}\times\frac{\sigma^2}{5} = \underline{\frac{\sigma^2}{25}}$$

(iv)

$RE\ (\bar{X}_M \text{ wrt } \bar{X}_W) = \dfrac{\left[V\left(\bar{X}_M\right)\right]^{-1}}{\left[V\left(\bar{X}_W\right)\right]^{-1}} = \dfrac{\left[\frac{\sigma^2}{16}\right]^{-1}}{\left[\frac{\sigma^2}{25}\right]^{-1}} = \underline{\dfrac{16}{25}}$ or $\underline{0.64}$ or $\underline{64\%}$

Prefer $\bar{X}_W$ since $V(\bar{X}_W) < V(\bar{X}_M)$

or $RE\ (\bar{X}_M \text{ wrt } \bar{X}_W) < 1$ or 100%

4.3 Question 3

(a)

E(X) = $\underline{np}$ V(X) = $\underline{np(1-p)}$

E(Y) = $\underline{2np}$ V(Y) = $\underline{2np(1-p)}$

(b)
(i)

$$\mathrm{E}(\hat{p}_1) = \frac{1}{2}\left[\frac{1}{n}\mathrm{E}(X)+\frac{1}{2n}\mathrm{E}(Y)\right] = \frac{1}{2}[p+p] = \underline{p} \quad \Rightarrow \quad \text{unbiased}$$

$$\mathrm{E}(\hat{p}_2) = \frac{1}{3n}[\mathrm{E}(X)+\mathrm{E}(Y)] = \frac{1}{3n}[np+2np] = \underline{p} \quad \Rightarrow \quad \text{unbiased}$$

(ii)

$$\mathrm{V}(\hat{p}_1) = \frac{1}{4}\left[\frac{1}{n^2}\mathrm{V}(X)+\frac{1}{4n^2}\mathrm{V}(Y)\right]$$

$$= \frac{1}{4}\left[\frac{np(1-p)}{n^2}+\frac{2np(1-p)}{4n^2}\right] = \frac{1}{4}\left[\frac{6p(1-p)}{4n}\right] = \underline{\frac{3p(1-p)}{8n}}$$

(iii)

$$\mathrm{V}(\hat{p}_2) = \frac{1}{9n^2}[\mathrm{V}(X)+\mathrm{V}(Y)]$$

$$= \frac{1}{9n^2}[np(1-p)+2np(1-p)] = \underline{\frac{p(1-p)}{3n}}$$

$$\mathrm{RE}\ (\hat{p}_1 \text{ wrt } \hat{p}_2) = \frac{[\mathrm{V}(\hat{p}_1)]^{-1}}{[\mathrm{V}(\hat{p}_2)]^{-1}} = \frac{\left[\frac{3p(1-p)}{8n}\right]^{-1}}{\left[\frac{p(1-p)}{3n}\right]^{-1}} = \underline{\frac{8}{9}} \text{ or } \underline{0.889} \text{ or } \underline{88.9\%}$$

4.4 Question 4

(a)

$$V(X) = \frac{(2\theta - 0)^2}{12} = \frac{4\theta^2}{12} = \underline{\frac{\theta^2}{3}}$$

(b)

$$E(\bar{X}) = \underline{\theta} \qquad V(\bar{X}) = \underline{\frac{\theta^2}{3n}}$$

(c)
(i)

$E(\bar{X}) = \theta \Rightarrow$ unbiased

(ii)

$V(\bar{X}) = \frac{\theta^2}{3n} \to 0$ as $n \to \infty$ $\Rightarrow$ consistent

(d)

$$\text{RE}(\bar{X} \text{ wrt } M) = \frac{[V(\bar{X})]^{-1}}{[V(M)]^{-1}} = \frac{\left[\frac{\theta^2}{3\times 5}\right]^{-1}}{\left[\frac{\theta^2}{7}\right]^{-1}} = \underline{\frac{15}{7}} \text{ or } \underline{2.14} \text{ or } \underline{214\%}$$

Prefer $\bar{X}$ since $V(\bar{X}) < V(M)$

or $\text{RE}(\bar{X} \text{ wrt } M) > 1$ or 100%

(e)
(i)

Largest sample value $\leq 2\theta$ $\Rightarrow$ $2.5 \leq 2\theta$ or $\underline{2\theta \geq 2.5}$

(ii)

$\bar{x} = \underline{1.16}$ $\qquad m = \underline{1.3}$

(iii)

From (i) $\theta \geq 1.25$

Now $m = 1.3 > 1.25$ but $\bar{x} = 1.16 < 1.25$

Thus choose m since:

$\bar{x}$ is an impossible value for θ

or

θ must be closer to m than to $\bar{x}$

or

$\bar{x}$ is less than 1.25