| Name | Date started | Target end date |
|---|---|---|
| | | |

GCE A LEVEL — FURTHER STATISTICS B QUESTION PACK

0985-01 (Legacy S3) · New spec A2 Unit 5 Topic 1

REVISE
.wales

# FURTHER MATHS — FS B · SAMPLES & POPULATIONS

## *Samples & Populations — Unbiased Estimators & the Variance Criterion*

*Every samples-and-populations / unbiased-estimator question from the legacy WJEC S3 papers (June 2006 – June 2017 + Specimen) that maps onto the new-spec A2 Unit 5 Topic 1.*

LEGACY 2008 SPECIFICATION

### Estimated time for entire question pack: ~4 hours 19 minutes

*Derived from the legacy S3 paper's pace of **~1.3 min/mark** (199 marks over 13 questions). The full Unit 5 exam is **1 hour 45 minutes for 80 marks** (25% of the A-level qualification, A2 optional paper alongside Unit 6 Further Mechanics B).*

*You are advised to **not** attempt to complete all of this in one sitting.*

### ABOUT THIS QUESTION PACK

This is a **comprehensive practice question pack**, not a single mock paper. It contains every samples & populations question from the legacy WJEC S3 papers (2008 modular spec) that maps onto new-spec A2 Unit 5 Topic 1 (2.5.1). Unit 5 (Further Statistics B) is one of two **80-mark A2 optional papers** (the other being Unit 6 Further Mechanics B), each worth 25% of the A-level qualification.

Questions are ordered roughly by topic / difficulty.

### INSTRUCTIONS

Use black ink or black ball-point pen. Show all working — method marks are awarded for clear setup.

*A calculator is allowed. The WJEC Statistical Tables and Formula Booklet may be referred to.*



*For Examiner's use only*

| Q | Source | Max | Mark |
|---|---|---|---|
| 1 | Jun 06 Q6 | 19 | |
| 2 | Jun 07 Q4 | 8 | |
| 3 | Jun 07 Q6 | 13 | |
| 4 | Jun 08 Q6 | 11 | |
| 5 | Jun 09 Q6 | 14 | |
| 6 | Jun 10 Q6 | 15 | |
| 7 | Jun 11 Q7 | 19 | |

| Q | Source | Max | Mark |
|---|---|---|---|
| 8 | Jun 12 Q6 | 20 | |
| 9 | Jun 13 Q7 | 17 | |
| 10 | Jun 14 Q6 | 14 | |
| 11 | Jun 15 Q6 | 16 | |
| 12 | Jun 16 Q6 | 13 | |
| 13 | Jun 17 Q7 | 20 | |
| | | | |
| | Total | 199 | |

# Samples & Populations — Unbiased Estimators & the Variance Criterion — what the new spec asks

WJEC GCE A Level Further Mathematics (from 2017) · Unit 5: Further Statistics B · Topic 2.5.1.

## Unbiased estimators 2.5.1

- A statistic $T$ is an *unbiased estimator* of a parameter $\theta$ if $E[T] = \theta$.
- The sample mean $\bar{X}$ is unbiased for the population mean $\mu$: $E[\bar{X}] = \mu$.
- The sample variance $s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$ is unbiased for $\sigma^2$.
- For probability/proportion $p$: $\hat{p} = X/n$ is unbiased, with $E[\hat{p}] = p$.

## Variance criterion 2.5.1

- Given two unbiased estimators $T_1, T_2$ for the same $\theta$, prefer the one with the smaller variance.
- The estimator with smaller variance gives, on average, more precise estimates.
- Used to pick the better of two linear combinations of the data, or to choose $\lambda$ in $\lambda\bar{X} + (1-\lambda)\bar{Y}$.
- Differentiate Var with respect to the weighting parameter and set $= 0$ to minimise.

## Standard error of $\bar{X}$ 2.5.1

- The *standard error* is the standard deviation of an estimator's sampling distribution.
- For the sample mean: $\mathrm{SE}(\bar{X}) = \dfrac{\sigma}{\sqrt{n}}$.
- If $\sigma$ is unknown, estimate it: $\widehat{\mathrm{SE}}(\bar{X}) = s/\sqrt{n}$.
- For a sample proportion: $\mathrm{SE}(\hat{p}) = \sqrt{\dfrac{p(1-p)}{n}}$; estimate with $\sqrt{\hat{p}(1-\hat{p})/n}$.

## Working scientifically general

- Always justify *why* an estimator is unbiased (compute $E[T]$ and show it equals $\theta$).
- When constructing an estimator $T = aX + b$, solve for $a, b$ using $E[T] = \theta$.
- Don't forget: if $T$ is unbiased for $\theta$, then $T^2$ is generally *not* unbiased for $\theta^2$.
- Sampling distributions of $\bar{X}$ may be derived by exhaustive enumeration when $n$ is small.

# Samples & Populations in one page

Quick-reference notes — revisit before each section. Don't use during questions.

### Unbiased estimator

A statistic $T$ is **unbiased** for parameter $\theta$ if:

$$E[T] = \theta$$

Bias is $E[T] - \theta$; an unbiased estimator has zero bias.

### Sample mean $\bar{X}$

For $X_1, \ldots, X_n$ with $E[X_i] = \mu$:

$$E[\bar{X}] = \mu \text{ (unbiased)}$$

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

Standard error: $\mathrm{SE}(\bar{X}) = \sigma/\sqrt{n}$.

### Sample variance $s^2$

The unbiased estimator of population variance $\sigma^2$ is:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Computational form: $s^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right)$.

### Why $n-1$?

Using $n$ in the denominator gives a *biased* estimator that systematically underestimates $\sigma^2$.

Dividing by $n-1$ corrects for the loss of one degree of freedom from estimating $\bar{x}$.

Proof: $E\left[\sum (X_i - \bar{X})^2\right] = (n-1)\sigma^2$.

### Unbiased estimator of $p$

For $X \sim B(n, p)$ with $E[X] = np$:

$$\hat{p} = \frac{X}{n} \text{ is unbiased}$$

$\mathrm{Var}(\hat{p}) = \frac{p(1-p)}{n}$; SE: $\sqrt{\frac{p(1-p)}{n}}$.

Estimate SE by substituting $\hat{p}$ for $p$.

### Variance criterion

Given two unbiased estimators $T_1, T_2$ of $\theta$, the **better** one has smaller variance.

For $T = \lambda\bar{X} + (1-\lambda)\bar{Y}$, find $\lambda$ that minimises $\mathrm{Var}(T)$:

$\frac{d}{d\lambda}\mathrm{Var}(T) = 0$ gives $\lambda^* = \frac{\sigma_Y^2/n}{\sigma_X^2/m + \sigma_Y^2/n}$.

### Building estimators

To make $T = aX + b$ unbiased for $\theta$:

Set $E[T] = \theta$ and solve for the constants.

Example: $E[X] = 2\theta + 1 \Rightarrow T = (X - 1)/2$ is unbiased.

### Common pitfalls

- Forgetting $s^2$ uses $n-1$, not $n$ — dividing by $n$ gives biased estimator.
- Assuming $T^2$ is unbiased for $\theta^2$ when $T$ is unbiased for $\theta$ — usually false.
- Confusing standard deviation with standard error — SE is for an estimator, not data.
- Mixing up $\mathrm{Var}(X)$ and $\mathrm{Var}(\bar{X}) = \sigma^2/n$.

### Strategy

**1.** For "is $T$ unbiased?": compute $E[T]$ and check if equal to $\theta$.

**2.** For "construct an unbiased estimator": write $T = aX + b$ and solve $E[T] = \theta$.

**3.** For "which is better?": compute Var for each, pick smaller.

**4.** For SE: take $\sqrt{\mathrm{Var}(T)}$.

# SECTION T1

*Samples & Populations*

Questions 1–13 · 199 marks

**6.** The random variable $X$ has mean $\mu$ and variance $\sigma^2$. Independently, the random variable $Y$ has mean $2\mu$ and variance $3\sigma^2$. Consider the following estimators for $\mu$.

$$U = a(X + 2Y); \quad V = b(2X + Y)$$

*(a)* Find the values of the constants $a$ and $b$ for which $U$ and $V$ are unbiased estimators for $\mu$. [5]

*(b)* Find the variances of these unbiased estimators and hence determine which of the two is the better estimator. [5]

*(c)* Now consider the estimator

$$W = \frac{X + kY}{1 + 2k} \quad \text{(where } k \neq -\tfrac{1}{2}\text{)}.$$

(i) Show that $W$ is an unbiased estimator for $\mu$ for all possible values of $k$.

(ii) Find the variance of $W$ in terms of $k$ and $\sigma$.

(iii) Hence find the value of $k$ which gives the best estimator of this form. [9]

[You may assume that any stationary value that you find is a minimum.]

**4.** The random variable $X$ has a Poisson distribution with unknown mean $\mu$. In order to estimate $\mu$, 100 observations are made on $X$ and it is found that $\sum x = 256$.

*(a)* Calculate an unbiased estimate of $\mu$. Deduce an unbiased estimate of the variance of $X$. [2]

*(b)* Estimate the standard error of your estimate. [2]

*(c)* Determine an approximate 90% confidence interval for $\mu$. [3]

*(d)* Explain briefly where your solution made use of the Central Limit Theorem. [1]

**6.** The table below shows the probability distribution of the discrete random variable $X$, where $\theta$ is an unknown parameter whose value lies between 0 and 0·25.

| $x$ | 1 | 2 | 3 |
|---|---|---|---|
| $P(X = x)$ | $3\theta$ | $1 - 4\theta$ | $\theta$ |

To estimate the value of $\theta$, $n$ independent observations are made on $X$. Let $X_1$ denote the number of occurrences of the value 1 and $X_2$ the number of occurrences of the value 2. Consider the following two estimators for $\theta$.

$$U_1 = \frac{X_1}{3n}, \quad U_2 = \frac{n - X_2}{4n}$$

*(a)* Show that both estimators are unbiased. [6]

*(b)* (i) Show that

$$\frac{\text{Var}(U_1)}{\text{Var}(U_2)} = \frac{4(1 - 3\theta)}{3(1 - 4\theta)}.$$

(ii) State, with a reason, which is the better estimator. [7]

**6.** The continuous random variable $X$ has probability density function $f$ where

$$f(x) = 1 + \lambda x, \qquad \text{for } -\frac{1}{2} \leqslant x \leqslant \frac{1}{2},$$

$$f(x) = 0, \qquad \text{otherwise,}$$

and $\lambda$ is an unknown parameter whose value lies between –2 and 2.

*(a)* Show that

$$P(X > 0) = \frac{1}{2} + \frac{\lambda}{8}.$$ [2]

*(b)* In order to estimate the value of $\lambda$, $n$ independent observations are made on $X$. Let $Y$ denote the number of these observations which are positive.

(i) Show that

$$U = \frac{8Y}{n} - 4$$

is an unbiased estimator for $\lambda$.

(ii) Obtain an expression, in terms of $n$ and $\lambda$, for the standard error of $U$. [9]

**6.** The independent random variables $X$ and $Y$ have a common mean $\mu$ and variances $\sigma_x^2$ and respectively. In order to estimate $\mu$, random samples of $m$ values of $X$ and $n$ values of $Y$ are taken. The means of these samples are denoted by $\overline{X}$ and $\overline{Y}$ respectively.

*(a)* Show that

$$U = \lambda \overline{X} + (1 - \lambda)\overline{Y}$$

is an unbiased estimator for $\mu$ for all values of the constant $\lambda$. [2]

*(b)* Find an expression for the variance of $U$. [3]

*(c)* (i) Determine the value of $\lambda$ which gives the best estimator for $\mu$.

(ii) Show that the standard error of the best estimator is

$$\frac{\sigma_x \sigma_y}{\sqrt{m\sigma_y^2 + n\sigma_x^2}} .$$ [9]

**6.** The probability distribution of the discrete random variable $X$ is given in the following table, where $0 < \theta < \frac{1}{3}$.

| $x$ | $-1$ | $0$ | $1$ |
|---|---|---|---|
| $P(X = x)$ | $\theta$ | $2\theta$ | $1 - 3\theta$ |

*(a)* Obtain an expression for $E(X)$ and show that

$$\text{Var}(X) = 2\theta(3 - 8\theta).$$ [3]

In order to estimate $\theta$, a random sample of $n$ observations of $X$ is taken.

*(b)* The mean of the observations in the sample is denoted by $\overline{X}$. Show that

$$U = \frac{1 - \overline{X}}{4}$$

is an unbiased estimator for $\theta$ and obtain an expression for the variance of $U$. [4]

*(c)* The number of observations in the sample equal to zero is denoted by $N$. Show that

$$V = \frac{N}{2n}$$

is an unbiased estimator for $\theta$ and obtain an expression for the variance of $V$. [5]

*(d)* Show that

$$\text{Var}(V) - \text{Var}(U) > 0$$

State, with a reason, which is the better estimator, $U$ or $V$. [3]

# TURN OVER

**7.** The probability density function of the continuous random variable $X$ is given by

$$f(x) = \frac{1}{2} + \theta x \qquad -1 \leqslant x \leqslant 1,$$
$$f(x) = 0 \qquad \text{otherwise,}$$

where $\theta$ is an unknown constant whose value lies between $-\frac{1}{2}$ and $\frac{1}{2}$.

*(a)* (i) Obtain an expression for $E(X)$ and show that

$$\text{Var}(X) = \frac{3 - 4\theta^2}{9}.$$

(ii) Show that

$$P(X > 0) = \left(\frac{1+\theta}{2}\right).$$ [8]

In order to estimate $\theta$, a random sample of $n$ observations of $X$ is taken.

*(b)* The mean of the observations in the sample is denoted by $\overline{X}$. Show that

$$U = \frac{3\overline{X}}{2}$$

is an unbiased estimator for $\theta$ and obtain an expression for the variance of $U$. [4]

*(c)* Let $Y$ denote the number of observations in the sample that are greater than zero. Show that

$$V = \frac{2Y}{n} - 1$$

is an unbiased estimator for $\theta$ and obtain an expression for the variance of $V$. [5]

*(d)* Show that

$$\text{Var}(V) - \text{Var}(U) = \frac{1}{4n}.$$

State, with a reason, which is the better estimator, $U$ or $V$. [2]

**6.** The probability density function of the continuous random variable $X$ is given by

$$f(x) = \frac{2x}{a^2} \qquad \text{for } 0 \leqslant x \leqslant a,$$

$$f(x) = 0 \qquad \text{otherwise,}$$

where $a$ is an unknown positive constant.

*(a)* Obtain an expression for $E(X)$ and show that

$$\text{Var}(X) = \frac{a^2}{18}.$$

[7]

*(b)* In order to estimate $a$, a random sample of $n$ observations of $X$ is taken.

(i) The mean of the observations in the sample is denoted by $\overline{X}$. Find the value of the constant $c$ such that

$$U = c\overline{X}$$

is an unbiased estimator for $a$ and obtain an expression for the variance of $U$.

(ii) Let $Y$ denote the largest observation in the sample. You are given that

$$E(Y) = \frac{2na}{2n+1} \quad \text{and} \quad \text{Var}(Y) = \frac{na^2}{(n+1)(2n+1)^2}.$$

Find the value of the constant $d$ such that

$$V = dY$$

is an unbiased estimator for $a$ and obtain an expression for the variance of $V$.

(iii) Show that

$$\frac{\text{Var}(U)}{\text{Var}(V)} = \frac{n+1}{2}.$$

State, with a reason, which is the better estimator, $U$ or $V$. [13]

**7.** The random variable $X$ has the binomial distribution $B(n, p)$.

*(a)* Show that

$$\hat{p} = \frac{X}{n}$$

is an unbiased estimator for $p$, and find its standard error. [4]

*(b)* (i) Show that $\hat{p}^2$ is not an unbiased estimator for $p^2$.

(ii) By considering $E[X(X-1)]$, find an unbiased estimator for $p^2$. [8]

*(c)* Given that $q = 1 - p$,

(i) deduce an unbiased estimator for $q^2$,

(ii) find an unbiased estimator for $pq$, simplifying your answer as far as possible. [5]

**6.** The continuous random variable $X$ is uniformly distributed on the interval $[0, \theta]$, where $\theta$ is unknown. In order to estimate $\theta$, a random sample of $n$ observations on $X$ is obtained and $\overline{X}$ denotes the mean of this sample. An unbiased estimator for $\theta$ is given by $Y = k\overline{X}$.

*(a)* (i) Find the value of $k$.

(ii) Find the standard error of $Y$. [8]

*(b)* (i) Show that $Y^2$ is not an unbiased estimator for $\theta^2$.

(ii) Find an unbiased estimator for $\theta^2$. [6]

**END OF PAPER**

**6.** The discrete random variable $X$ has the following probability distribution.

| $x$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P(X = x)$ | $\theta$ | $2\theta$ | $3\theta$ | $1 - 6\theta$ |

where $\theta$ is an unknown constant, $0 < \theta < \frac{1}{6}$. In order to estimate $\theta$, a random sample of $n$ observations on $X$ is obtained and $\bar{X}$ denotes the sample mean.

*(a)* Given that $U = a\bar{X} + b$ is an unbiased estimator for $\theta$, determine

(i) the constants $a$ and $b$,

(ii) the variance of $U$. [9]

*(b)* Let $Y$ denote the number of observations in the sample equal to 4. Given that $V = cY + d$ is an unbiased estimator for $\theta$, determine

(i) the constants $c$ and $d$,

(ii) the variance of $V$. [5]

*(c)* Show that

$$\frac{\mathrm{Var}(U)}{\mathrm{Var}(V)} = \frac{6 - 30\theta}{5 - 30\theta}.$$

Hence, giving a reason, state which of $U$ and $V$ is the better estimator for $\theta$. [2]

**END OF PAPER**

**6.** A random sample $X_1, X_2, \ldots, X_n$ is taken from a probability distribution with mean $\mu$ and variance $\sigma^2$. The sample mean is denoted by $\overline{X}$.

*(a)* (i) Show that $\overline{X}$ is an unbiased estimator for $\mu$.

(ii) Show that the standard error of $\overline{X}$ is $\dfrac{\sigma}{\sqrt{n}}$. [4]

*(b)* (i) Show that

$$E\left(X_i^{\,2}\right) = \mu^2 + \sigma^2 .$$

(ii) Given that

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

show that $S^2$ is an unbiased estimator for $\sigma^2$. [5]

*(c)* By considering the variance of $S$, show that $S$ is not an unbiased estimator for $\sigma$. [4]

**END OF PAPER**

**7.** An electronic device generates random digits from the set {1, 2, 3, 4}. The probability distribution of the digit generated, $X$, is given by

$$P(X = x) = \begin{cases} p & \text{for } x = 1 \\ \frac{(1-p)}{3} & \text{for } x = 2, 3, 4 \end{cases}$$

where $p$ is an unknown constant, $0 < p < 1$.

*(a)* (i) Determine an expression for $E(X)$ in terms of $p$.

(ii) Show that

$$\text{Var}(X) = \frac{2}{3}(1-p)(1+6p).$$

[7]

*(b)* In order to estimate $p$, a random sample of $n$ digits is generated using the device and $\bar{X}$ denotes the sample mean.

(i) Show that

$$U = \frac{3-\bar{X}}{2}$$

is an unbiased estimator for $p$.

(ii) Determine an expression for Var($U$) in terms of $n$ and $p$. [4]

*(c)* The number of digits in the random sample equal to 1 is denoted by $Y$.

(i) Write down the distribution of $Y$.

(ii) Show that

$$V = \frac{Y}{n}$$

is an unbiased estimator for $p$.

(iii) Determine an expression for Var($V$) in terms of $n$ and $p$. [5]

*(d)* By considering $\frac{\text{Var}(U)}{\text{Var}(V)}$, determine which is the better estimator, $U$ or $V$. [4]

**END OF PAPER**

## END OF SAMPLES & POPULATIONS PACK

Source: WJEC S3 (2008 modular spec) · 2005–2017

Curated for WJEC FM 2017 spec A2 Unit 5 — Topic 1 (2.5.1)