

**KEY WORDS** sampling distribution • statistic

## Objective

## 1 Key knowledge and worked examples

**Sampling words.** A census measures the whole population: complete, no sampling error, but slow, costly, possibly destructive. A sample survey is cheap and fast but carries sampling error. The sampling frame lists sampling units. A statistic is a function of the sample values **ONLY** (contains no unknown parameter); its distribution across repeated samples is the sampling distribution.

**Worked example – a score distribution (WST02 June 2025 Q4).** Three beads from a 2:5 black:white bag score 4 per black, 1 per white: with  $p_b = \frac{2}{7}$ ,  $X$  takes 3, 6, 9, 12 with probabilities  $\frac{125}{343}, \frac{150}{343}, \frac{60}{343}, \frac{8}{343}$  (binomial weights for 0, 1, 2, 3 blacks). The number  $Y$  of 9-point sets in  $n$  plays is  $B(n, \frac{60}{343})$ , and  $1 - \left(\frac{283}{343}\right)^n > 0.95$  needs  $n = 16$ .

**2.1** State one advantage and one disadvantage of a sample survey compared with a census. [2]

**2.2** For samples  $(X_1, X_2)$  from a population with mean  $\mu$ , state which of  $X_1 + 2X_2$ ,  $\bar{X} - \mu$  and  $3\bar{X}$  are statistics, with reasons. [3]

**2.3** A fair coin is tossed twice;  $S$  = number of heads. Write out the sampling distribution of  $S$ . [2]

### 3 Exam-style questions

**3.1** (*Edexcel IAL WST02 June 2025 Q4.*) A bag contains a large number of beads of which the ratio of black to white beads is 2:5. A player takes 3 beads at random from the bag and scores 4 points for each black bead and 1 point for each white bead. The random variable  $X$  is the total number of points scored. (a) Find the sampling distribution of  $X$ . (b) In a game a player takes  $n$  sets of 3 beads (each set replaced before the next);  $Y$  is the number of sets scoring exactly 9 points. Find the least  $n$  such that  $P(Y > 1) > 0.95$ . [9]

**3.2** (*Original.*) Two fair dice are rolled;  $M$  is the larger score shown. (a) Find the sampling distribution of  $M$ . (b) Find  $E(M)$ . [7]

---

---

---

---

---

---

---

---

## Solutions

---

**2.1** Advantage: cheaper and faster than a census. Disadvantage: the results contain sampling error – they do not describe every member of the population.

**2.2**  $X_1 + 2X_2$  and  $3\bar{X}$  involve sample values only: statistics.  $\bar{X} - \mu$  contains the unknown population parameter  $\mu$ : not a statistic.

**2.3**  $P(S = 0) = \frac{1}{4}$ ,  $P(S = 1) = \frac{1}{2}$ ,  $P(S = 2) = \frac{1}{4}$ .

**3.1** (a) With  $p_b = \frac{2}{7}$  (large bag: constant probability), the number of black beads is  $B(3, \frac{2}{7})$  and  $X = 3 + 3 \cdot (\text{blacks})$ :  $P(X = 3) = (\frac{5}{7})^3 = \frac{125}{343}$ ;  $P(X = 6) = \binom{3}{1} \frac{2}{7} (\frac{5}{7})^2 = \frac{150}{343}$ ;  $P(X = 9) = \binom{3}{2} (\frac{2}{7})^2 \frac{5}{7} = \frac{60}{343}$ ;  $P(X = 12) = (\frac{2}{7})^3 = \frac{8}{343}$  (total  $\frac{343}{343}$  ✓). (b)  $Y \sim B(n, \frac{60}{343})$ :  $P(Y \geq 1) = 1 - (\frac{283}{343})^n > 0.95$ , so  $(\frac{283}{343})^n < 0.05$ ,  $n \ln \frac{283}{343} < \ln 0.05$ ,  $n > 15.6$ : least  $n = 16$ .

**3.2** (a)  $P(M = m) = P(\max = m) = \left(\frac{m}{6}\right)^2 - \left(\frac{m-1}{6}\right)^2 = \frac{2m-1}{36}$  for  $m = 1..6$ :  $\frac{1}{36}, \frac{3}{36}, \frac{5}{36}, \frac{7}{36}, \frac{9}{36}, \frac{11}{36}$  (sums to 1 ✓). (b)  $E(M) = \sum m \cdot \frac{2m-1}{36}$ : summing  $m(2m-1) = 2m^2 - m$  over  $m = 1..6$  gives  $2(91) - 21 = 161$ , so  $E(M) = \frac{161}{36} = 4.47$ .

Local review guidance: enumeration with the correct per-draw probability and binomial weights; distribution sums to 1; follow-up counts wrapped in  $B(n, p)$  with the complement taken before logs. This is authored classroom guidance.