Estimation, confidence intervals and tests using a normal distribution 5A

1 i a $\sum X_i \sim \mathrm{N}(10\mu, 10\sigma^2)$

b $\dfrac{2X_1 + 3X_{10}}{5} \sim \mathrm{N}\left(\mu, \dfrac{13}{25}\sigma^2\right), \left(\dfrac{13}{25} = \dfrac{2^2+3^2}{5^2}\right)$

c $\mathrm{E}(X_i - \mu) = 0 \quad \mathrm{Var}(X_I - \mu) = \mathrm{Var}(X_i) = \sigma^2$

$\therefore \sum (X_i - \mu) \sim \mathrm{N}(0, 10\sigma^2)$

d $\overline{X} = \dfrac{\sum X_i}{n} \sim \mathrm{N}\left(\mu, \dfrac{\sigma^2}{10}\right) \quad (n = 10)$

e $\sum_{i=1}^{5} X_i - \sum_{i=6}^{10} X_i \sim \mathrm{N}(5\mu, 5\sigma^2) - \mathrm{N}(5\mu, 5\sigma^2)$

$\therefore$ combined distribution $\sim \mathrm{N}(0, 10\sigma^2)$

[Remember $\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$]

f $\dfrac{X_I - \mu}{\sigma} \sim \mathrm{N}(0, 1^2) \ \therefore \sum\left(\dfrac{X_i - \mu}{\sigma}\right) \sim \mathrm{N}(0, 10)$

ii **a**, **b**, **d**, **e** are statistics since they do not contain μ or σ, the unknown population parameters.

2 a X = value of a coin.

x	1	5	10
$\mathrm{P}(X = x)$	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{1}{5}$

$\therefore \mu = \mathrm{E}(X) = \dfrac{2}{5} + \dfrac{10}{5} + \dfrac{10}{5} = \dfrac{22}{5}$ or 4.4

$\mathrm{E}(X^2) = 1^2 \times \dfrac{2}{5} + 25 \times \dfrac{2}{5} + 100 \times \dfrac{1}{5} = \dfrac{152}{5}$

$\therefore \sigma^2 = \mathrm{E}(X^2) - \mu^2 = \dfrac{152}{5} - \dfrac{22^2}{25} = 11.04$ or $\dfrac{276}{25}$

b $\{1, 1\} \quad \{1, 5\}^{\times 2} \quad \{1, 10\}^{\times 2}$

$\{5, 5\} \quad \{5, 10\}^{\times 2}$

$\{10, 10\}$

2 c

$\bar{x}$	1	3	5	5.5	7.5	10
$P(\bar{X}=\bar{x})$	$\frac{4}{25}$	$\frac{8}{25}$	$\frac{4}{25}$	$\frac{4}{25}$	$\frac{4}{25}$	$\frac{1}{25}$

$$\left[\text{e.g. } P(\bar{X}=5.5) = \frac{2}{5}\times\frac{1}{5}+\frac{1}{5}\times\frac{2}{5}=\frac{4}{25}\right]$$

d $E(\bar{X}) = 1\times\frac{4}{25}+3\times\frac{8}{25}+\cdots+10\times\frac{1}{25}=4.4=\mu$

$\text{Var}(\bar{X}) = 1^2\times\frac{4}{25}+3^2\times\frac{8}{25}+\cdots+10^2\times\frac{1}{25}-4.4^2=5.52=\frac{\sigma^2}{2}$

3 Unbiased estimate of mean $=\bar{x}=\frac{\sum x}{n}$

Unbiased estimate of variance $=S^2=\frac{\sum x^2-n\bar{x}^2}{n-1}$

a $\sum x = 270.3,\ \sum x^2 = 5270.49,\ n=14$

$\therefore\ \bar{x}=19.3,\ S^2=3.98$

b $\sum x = 54,\ \sum x^2 = 252,\ n=16$

$\therefore\ \bar{x}=3.375,\ S^2=4.65$

c $\sum x = 2007.4,\ \sum x^2 = 505132.36,\ n=9$

$\therefore\ \bar{x}=223,\ S^2=7174$

d $\sum x = 5.833,\ \sum x^2 = 3.644555,\ n=10$

$\therefore\ \bar{x}=0.5833,\ S^2=0.0269$

4 a $\bar{x}=\frac{\sum x}{n}=\frac{4368}{120}=36.4$

$S^2=\frac{\sum x^2-n\bar{x}^2}{n-1}=\frac{162.466-120\times36.4^2}{119}=29.166\ldots$

$=29.2\ (3\text{s.f.})$

b $\bar{x}=\frac{\sum x}{n}=\frac{270}{30}=9$

$S^2=\frac{\sum x^2-n\bar{x}^2}{n-1}=\frac{2546-30\times9^2}{29}=4$

4 c $\bar{x} = \frac{\sum x}{n} = \frac{1140.7}{1037} = 1.1$

$S^2 = \frac{\sum x^2 - n\bar{x}^2}{n-1} = \frac{1278.08 - 1037 \times 1.1^2}{1036} = 0.0225$

d $\bar{x} = \frac{\sum x}{n} = \frac{168}{15} = 11.2$

$S^2 = \frac{\sum x^2 - n\bar{x}^2}{n-1} = \frac{1913 - 15 \times 11.2^2}{14} = 2.24285\ldots$

$= 2.24$ (3 s.f.)

5 a An unbiased estimator is an estimator of a population parameter that will 'on average' give the correct value.

b $\sum x = 1652, \sum x^2 = 389\ 917.48,\ n = 7$

$\therefore\ \bar{x} = \frac{1652}{7} = 236$

$S^2 = \frac{389\ 917.48 - 7 \times 236^2}{6}$

$= 7.58$

6 $\sum x = 2051.6, \sum x^2 = 420\ 989.26,\ n = 10$

$\bar{x} = 205.16 = 205$ (3 s.f.)

$S^2 = \frac{420\ 989.26 - 10 \times \bar{x}^2}{9}$

$= 9.22266\ldots = 9.22$ (3 s.f.)

7 X = length of a bolt

x	5	10
$\mathrm{P}(X = x)$	$\frac{2}{3}$	$\frac{1}{3}$

a $\mu = 5 \times \frac{2}{3} + 10 \times \frac{1}{3} = \frac{20}{3}$

$\sigma^2 = 25 \times \frac{2}{3} + 100 \times \frac{1}{3} - \left(\frac{20}{3}\right)^2 = \frac{50}{9}$

b $\{5, 5, 5\}$

$\{5, 5, 10\}^{\times 3}$ $\quad \{5, 10, 10\}^{\times 3}$

$\{10, 10, 10\}$

7 c

$\bar{x}$	5	$\frac{20}{3}$	$\frac{25}{3}$	10
$P(\bar{X}=\bar{x})$	$\frac{8}{27}$	$\frac{12}{27}$	$\frac{6}{27}$	$\frac{1}{27}$

d $E(\bar{X}) = 5\times\frac{8}{27} + \frac{20}{3}\times\frac{12}{27} + \cdots + 10\times\frac{1}{27} = \frac{20}{3} = \mu$

$\text{Var}(\bar{X}) = 5^2\times\frac{8}{27} + \cdots + 10^2\times\frac{1}{27} - \left(\frac{20}{3}\right)^2 = \frac{50}{27} = \frac{\sigma^2}{3}$

e

m	5	10
$P(M=m)$	$\frac{20}{27}$	$\frac{7}{27}$

P(M = 10) is cases {5, 10, 10} and {10, 10, 10}

f $E(M) = 5\times\frac{20}{27} + 10\times\frac{7}{27} = \frac{170}{27} = 6.296\ldots$

$\text{Var}(M) = 25\times\frac{20}{27} + 100\times\frac{7}{27} - \left(\frac{170}{27}\right)^2 = \frac{3500}{729} = 4.80\ldots$

g $\text{Bias} = E(M) - 5 = 1.296\ldots = 1.30$ (3 s.f.)

8 $X \sim B(10, p)$

a $E(X) = np = 10p$

b $\bar{X} = \frac{X_1 + \ldots + X_{25}}{25}$

$E(\bar{X}) = \frac{E(X_1) + E(X_2) + \ldots + E(X_{25})}{25} = \frac{\mu + \mu + \ldots + \mu}{25} = \frac{25\mu}{25} = \mu$

$\therefore \bar{X}$ is an unbiased estimator of μ

But $E(\bar{X}) = \frac{E(X_1) + E(X_2) + \ldots + E(X_{25})}{25} = \frac{25\times 10p}{25} = 10p$

$\therefore$ $\bar{X}$ is a *biased* estimator of p.

so bias = $10p - p = 9p$

c $E\left(\frac{\bar{X}}{10}\right) = \frac{1}{10}E(\bar{X}) = p$

$\therefore \frac{\bar{X}}{10}$ is an unbiased estimator of p.

9 $X \sim \text{U}[-\alpha, \alpha]$

a $\text{E}(X) = \dfrac{-\alpha + \alpha}{2} = 0$

$\text{Var}(X) = \dfrac{(\alpha - (-\alpha))^2}{12} = \dfrac{4a^2}{12} = \dfrac{\alpha^2}{3}$

> μ and σ^2 for U[a, b] are given in the formula booklet

$\therefore \text{E}(X^2) = \text{Var}(X) + [\text{E}(X)]^2$

$\therefore \text{E}(X^2) = \dfrac{\alpha^2}{3} + 0 = \dfrac{\alpha^2}{3}$

b $Y = X_1^2 + X_2^2 + X_3^2$

$\text{E}(Y) = \text{E}\left(X_1^2\right) + \text{E}\left(X_2^2\right) + \text{E}\left(X_3^2\right)$

$= \dfrac{\alpha^2}{3} \times 3 = \alpha^2$

$\therefore$ Y is an unbiased estimator of α^2.

10 a $\bar{y} = \dfrac{486}{30} = 16.2$

$S_y^2 = \dfrac{8222 - 30 \times 16.2^2}{29} = 12.0275\ldots = 12.0 \text{ (3 s.f.)}$

b Let $\sum w = \sum x + \sum y$

$\bar{x} = 15.5 \Rightarrow \sum x = 15.5 \times 20 = 310$

$\therefore \sum w = 796$

$S_x^2 = 8.0 \Rightarrow \sum x^2 = 8 \times 19 + 20 \times 15.5^2 = 4957$

$\therefore \sum w^2 = 13179$

$\therefore \bar{w} = \dfrac{796}{50} = 15.92$

$S_w^2 = \dfrac{13179 - 50 \times 15.92^2}{49} = 10.340\ldots = 10.34$

c Standard error is a measure of the statistical accuracy of an estimate.

d Standard error of the mean is $\dfrac{S}{\sqrt{n}}$

$\dfrac{S_x}{\sqrt{20}} = 0.632 \text{ (3 s.f.)}, \dfrac{S_y}{\sqrt{30}} = 0.633 \text{ (3 s.f.)}, \dfrac{S_w}{\sqrt{50}} = 0.455 \text{(3s.f.)}$

e Prefer to use $\bar{w}$ since it is based on a larger sample size and has smallest standard error.

11 a $\bar{x} = \frac{1300}{20} = 65$ $S_x^2 = \frac{84\,685 - 20 \times 65^2}{19}$

$= 9.7368\ldots = 9.74$ (3 s.f.)

b To achieve a standard error less than 0.5, we require $\frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{n}} < 0.5.$

Rearranging the right hand side of the above formula, followed by squaring both sides gives an expression for n

$\frac{3}{\sqrt{n}} < 0.5$

$\sqrt{n} > 6$

$n > 36$

thus we need a sample of 37 or more.

c No. Because the recommendation is based on the assumed value of s^2 from the original sample OR the value of s^2 for the new sample might be different / larger.

d Let y = combined sample

$\sum y = 1300 + 1060$

$\sum y = 2360$ $n = 36$

$\therefore \bar{y} = \frac{2360}{36} = 65.555\ldots$ or 65.6 (3 s.f.)

12 $\frac{\sigma}{\sqrt{n}} < 0.5$

$\sigma = 2.6 \Rightarrow \sqrt{n} > 2.6 \times 2 = 5.2$

i.e. $n > 27.04$

So need a sample of 28 (or more)

13 a Let x = indentation

$\sum x = 48.9, \ \sum x^2 = 239.89, \ n = 10$

$\hat{\mu} = \bar{x} = \frac{48.9}{10} = 4.89$

b $\hat{\sigma}^2 = S^2 = \frac{239.89 - 10 \times 4.89^2}{9} = 0.08544\ldots$

$\frac{S}{\sqrt{n}} = 0.092436\ldots = 0.0924$ (3 s.f.)

13 c Require $\frac{S}{\sqrt{n}} = \frac{0.2923\ldots}{\sqrt{n}} < 0.05$

$\Rightarrow \sqrt{n} > 5.846\ldots$

$n > 34.17\ldots$

$\therefore$ need $n = 35$ (or more)

14 $X_1 \sim \mathrm{B}(n,\ p) \qquad X_2 \sim \mathrm{B}(2n,\ p)$

a $\mathrm{E}(X_1) = np,\ \mathrm{E}(X_2) = 2np,\ \mathrm{Var}(X_1) = np(1-p),\ \mathrm{Var}(X_2) = 2np(1-p)$

b $\mathrm{E}\left(\frac{X_1}{n}\right) = \frac{\mathrm{E}(X_1)}{n} = \frac{np}{n} = p \ \therefore\ \frac{X_1}{n}$ is unbiased estimator of p

$\mathrm{E}\left(\frac{X_2}{2n}\right) = \frac{\mathrm{E}(X_2)}{2n} = \frac{2np}{2n} = p \ \therefore\ \frac{X_2}{2n}$ is unbiased estimator of p

Prefer $\frac{X_2}{2n}$ since based on a larger sample (and therefore will have smaller variance)

c $X = \frac{1}{2}\left(\frac{X_1}{n} + \frac{X_2}{2n}\right) \Rightarrow \mathrm{E}(X) = \frac{1}{2}\left[\frac{\mathrm{E}(X_1)}{n} + \frac{\mathrm{E}(X_2)}{2n}\right]$

$$= \frac{1}{2}\left[\frac{np}{n} + \frac{2np}{2n}\right]$$

$$= \frac{1}{2}\left[p + p\right] = p$$

$\therefore$ X is an unbiased estimator of p

d $Y = \left(\frac{X_1 + X_2}{3n}\right) \Rightarrow \mathrm{E}(Y) = \frac{\mathrm{E}(X_1) + \mathrm{E}(X_2)}{3n} = \frac{np + 2np}{3n} = p$

$\therefore$ Y is an unbiased estimator of p

e $\mathrm{Var}\left(\frac{X_1}{n}\right) = \frac{1}{n^2}\ \mathrm{Var}(X_1) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$

$\mathrm{Var}\left(\frac{X_2}{2n}\right) = \frac{1}{4n^2}\ \mathrm{Var}(X_2) = \frac{2np(1-p)}{4n^2} = \frac{p(1-p)}{2n}$

$\mathrm{Var}(X) = \frac{1}{4}\left[\mathrm{Var}\left(\frac{X_1}{n}\right) + \mathrm{Var}\left(\frac{X_2}{2n}\right)\right] = \frac{1}{4}\left[\frac{p(1-p)}{n} + \frac{p(1-p)}{2n}\right] = \frac{3p(1-p)}{8n}$

$\mathrm{Var}(Y) = \frac{1}{9n^2}\left[\mathrm{Var}(X_1) + \mathrm{Var}(X_2)\right] = \frac{1}{9n^2}\left[np(1-p) + 2np(1-p)\right]$

$\mathrm{Var}(Y) = \frac{3np(1-p)}{9n^2} = \frac{p(1-p)}{3n}$

$\therefore$ Var (Y) is smallest so Y is the best estimator.

14 f $T = \left(\frac{2X_1 + X_2}{3n}\right)$

$$E(T) = \frac{2E(X_1) + E(X_2)}{3n} = \frac{2np + 2np}{3n} = \frac{4p}{3}$$

$$\text{bias} = E(T) - p$$
$$= \frac{p}{3}$$

15 Let X = number on a counter.

x	0	1	2
$P(X = x)$	0.4	0.2	0.4

a $\mu = 1$ (by symmetry)

$$\sigma^2 = 0 + 1^2 \times 0.2 + 2^2 \times 0.4 - 1 = 0.8 \text{ or } \frac{4}{5}$$

b $\{0, 0, 0\}$ $\{0, 0, 1\}^{\times 3}$ $\{0, 0, 2\}^{\times 3}$
$\{1, 1, 1\}$ $\{1, 1, 0\}^{\times 3}$ $\{1, 1, 2\}^{\times 3}$
$\{2, 2, 2\}$ $\{2, 2, 0\}^{\times 3}$ $\{2, 2, 1\}^{\times 3}$ $\{0, 1, 2\}^{\times 3! = 6}$

c

$\bar{x}$	0	$\frac{1}{3}$	$\frac{2}{3}$	1	$\frac{4}{3}$	$\frac{5}{3}$	2
$P(\bar{X} = \bar{x})$	$\frac{8}{125}$	$\frac{12}{125}$	$\frac{30}{125}$	$\frac{25}{125}$	$\frac{30}{125}$	$\frac{12}{125}$	$\frac{8}{125}$

d $E(\bar{X}) = 1$ (by symmetry) $(= \mu)$

$$\text{Var}(\bar{X}) = 0 + \frac{1}{9} \times \frac{12}{125} + \frac{4}{9} \times \frac{30}{125} + \cdots + 4 \times \frac{8}{125} - 1^2$$
$$= \frac{4}{15} \qquad \left(= \frac{\sigma^2}{3}\right)$$

e

n	0	1	2
$P(N = n)$	$\frac{44}{125}$	$\frac{37}{125}$	$\frac{44}{125}$

e.g. $P(N = 2)$ is cases $\{2, 2, 2\}$; $\{2, 2, 0\}$; $\{2, 2, 1\}$

f $E(N) = 1$ (by symmetry)

$$\text{Var}(N) = 0 + 1^2 \times \frac{37}{125} + 2^2 \times \frac{44}{125} - 1^2 = \frac{88}{125} \quad (= \sigma^2)$$

g $\because E(N) = 1 = \mu \therefore N$ is an unbiased estimator of μ.

h $\because \text{Var}(\overline{X}) < \text{Var}(N)$ choose $\overline{X}$

Challenge

a Starting with the given equation, we find that

$$\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$$

$$=\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i^2 - 2x_i\overline{x} + \overline{x}^2\right)$$

Now we split the summation as well as noting that $\overline{x}$ is constant and so can be pulled out of the summation

$$\frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - 2\overline{x}\sum_{i=1}^{n}x_i + \overline{x}^2\sum_{i=1}^{n}1\right).$$

We now note that $\sum_{i=1}^{n}x_i = n\overline{x}$ and $\sum_{i=1}^{n}1 = n$ then substitute into the above expression to obtain

$$\frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - 2n\overline{x}^2 + n\overline{x}^2\right)$$

$$=\frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - n\overline{x}^2\right).$$

Challenge

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

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

from the result in part **a**.

We want to show that $E(S^2) = \sigma^2$.

$$E(S^2) = E\left(\frac{1}{n-1}\left(\sum_{i=1}^{n} X^2 - n\bar{X}^2\right)\right)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} E(X^2) - E(n\bar{X}^2)\right)$$

$$= \frac{1}{n-1}\left(nE(X^2) - nE(\bar{X}^2)\right).$$

We should recognise that $E(X^2)$ is part of the formula

$\text{Var}(X) = E(X^2) - E(X)^2$

which can be rewritten as

$\sigma^2 = E(X^2) - \mu^2$

and we can rearrange for a more useful expression

$E(X^2) = \sigma^2 + \mu^2$.

Now we deal with the $E(\bar{X}^2)$ in the same way and find that

$\text{Var}(\bar{X}) = E(\bar{X}^2) - E(\bar{X})^2$

which is rewritten

$\frac{\sigma^2}{n} = E(\bar{X}^2) - \mu^2$

and we now rearrange to get

$E(\bar{X}^2) = \frac{\sigma^2}{n} + \mu^2$.

Now that we have useful expressions for $E(X^2)$ and $E(\bar{X}^2)$, we substitute back into the original expression to obtain

$$E(S^2) = \frac{1}{n-1}\left(nE(X^2) - nE(\bar{X}^2)\right)$$

$$= \frac{1}{n-1}\left(n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right)$$

$$= \frac{1}{n-1}\left(n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2\right)$$

$$= \sigma^2.$$

Thus the statistic S^2 is an unbiased estimator of the population variance σ^2.