

# SL Paper 1

At an early stage in analysing the marks scored by candidates in an examination paper, the examining board takes a random sample of 250 candidates and finds that the marks,  $x$ , of these candidates give  $\sum x = 10985$  and  $\sum x^2 = 598736$ .

- a. Calculate a 90% confidence interval for the population mean mark  $\mu$  for this paper. [4]
- b. The null hypothesis  $\mu = 46.5$  is tested against the alternative hypothesis  $\mu < 46.5$  at the  $\lambda\%$  significance level. Determine the set of values of  $\lambda$  [4]  
for which the null hypothesis is rejected in favour of the alternative hypothesis.

## Markscheme

a.  $\bar{x} = 43.94$  (A1)

unbiased variance estimate = 466.0847 (A1)

**Note:** Accept sample variance = 464.2204.

$\Rightarrow$  90% confidence interval is (41.7, 46.2) A1A1

[4 marks]

b. Z-value is  $-1.87489$  or  $-1.87866$  (A1)

probability is 0.0304 or 0.0301 (A1)

$\Rightarrow \lambda \geq 3.01$  (M1)A1

[4 marks]

## Examiners report

a. [N/A]

b. [N/A]

A random sample  $X_1, X_2, \dots, X_n$  is taken from the normal distribution  $N(\mu, \sigma^2)$ , where the value of  $\mu$  is unknown but the value of  $\sigma^2$  is known.

The mean of the sample is denoted by  $\bar{X}$ .

A mathematics teacher, wishing to apply the above result, generates some artificial data, assumes a value for the variance and calculates the following 95% confidence interval for  $\mu$ ,

$$[3.12, 3.25].$$

The teacher asks Alun to interpret this result. Alun makes the following statement. "The value of  $\mu$  lies in the interval  $[3.12, 3.25]$  with probability 0.95."

a.i.State the distribution of  $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ . [1]

a.ii.Hence show that, with probability 0.95, [4]

$$\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}.$$

b.i.Explain briefly why this is an incorrect statement. [1]

b.ii.Give a correct interpretation. [1]

## Markscheme

a.i.  $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$  is  $N(0, 1)$  or it has the Z-distribution **A1**

[?? marks]

a.ii.attempt to make a probability statement **R1**

therefore with probability 0.95,

$$-1.96 \leq \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq 1.96 \quad \mathbf{A1}$$

$$-1.96 \frac{\sigma}{\sqrt{n}} \leq \bar{X} - \mu \leq 1.96 \frac{\sigma}{\sqrt{n}} \quad \mathbf{A1}$$

$$1.96 \frac{\sigma}{\sqrt{n}} \geq \mu - \bar{X} \geq -1.96 \frac{\sigma}{\sqrt{n}} \quad \mathbf{A1}$$

$$\bar{X} + 1.96 \frac{\sigma}{\sqrt{n}} \geq \mu \geq \bar{X} - 1.96 \frac{\sigma}{\sqrt{n}}$$

**Note:** Award the final **A1** for either of the above two lines.

$$\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}} \quad \mathbf{AG}$$

[?? marks]

b.i.you cannot make a probability statement about a constant lying in a constant interval **OR** the mean either lies in the interval or it does not **A1**

[1 mark]

b.ii.the confidence interval is the observed value of a random interval

**OR** if the process is carried out a large number of times,  $\mu$  will lie in the interval 95% of the times **A1**

[1 mark]

## Examiners report

a.i. [N/A]

a.ii. [N/A]

b.i. [N/A]

b.ii. [N/A]

The lifetime, in years, of a randomly chosen basic vacuum cleaner is assumed to be modelled by the normal distribution  $B \sim N(14, 3^2)$ .

The lifetime, in years, of a randomly chosen robust vacuum cleaner is assumed to be modelled by the normal distribution  $R \sim N(20, 4^2)$ .

- a. Find  $P\left(B > E(B) + \frac{1}{2}\sqrt{\text{Var}(B)}\right)$ . [2]
- b. Find the probability that the total lifetime of 7 randomly chosen basic vacuum cleaners is less than 100 years. [4]
- c. Find the probability that the total lifetime of 5 randomly chosen robust vacuum cleaners is greater than the total lifetime of 7 randomly chosen basic vacuum cleaners. [5]

## Markscheme

a.  $P(B > 15.5)$  ( $= P(Z > 0.5)$ ) (M1)

$$= (1 - 0.69146) = 0.309 \quad \mathbf{A1}$$

[2 marks]

b. consider  $V = B_1 + B_2 + B_3 + B_4 + B_5 + B_6 + B_7$  (M1)

$$E(V) = 98 \quad \mathbf{A1}$$

$$\text{Var}(V) = 63 \text{ or equivalent} \quad \mathbf{A1}$$

**Note:** No need to state  $V$  is normal.

$$P(V < 100) = \left(P\left(Z < \frac{2}{\sqrt{63}} = 0.251976\dots\right)\right) = 0.599 \quad \mathbf{A1}$$

[4 marks]

c. consider  $W = R_1 + R_2 + R_3 + R_4 + R_5 - (B_1 + B_2 + B_3 + B_4 + B_5 + B_6 + B_7)$  (M1)

$$E(W) = 2 \quad \mathbf{A1}$$

$$\text{Var}(W) = 80 + 63 = 143 \quad \mathbf{A1}$$

$$P(W > 0) = \left(P\left(Z < \frac{2}{\sqrt{143}}\right)\right) \quad \mathbf{M1}$$

$$= 0.566 \quad \mathbf{A1}$$

[5 marks]

## Examiners report

- a. This was one of the more successful questions on the paper with many wholly correct answers seen. Only a very small number failed to complete part (a) successfully. There were also many fully correct answers to part (b). Part (c) caused a problem for some candidates where in most of those cases they failed to calculate the variance correctly.

- b. This was one of the more successful questions on the paper with many wholly correct answers seen. Only a very small number failed to complete part (a) successfully. There were also many fully correct answers to part (b). Part (c) caused a problem for some candidates where in most of those cases they failed to calculate the variance correctly.
- c. This was one of the more successful questions on the paper with many wholly correct answers seen. Only a very small number failed to complete part (a) successfully. There were also many fully correct answers to part (b). Part (c) caused a problem for some candidates where in most of those cases they failed to calculate the variance correctly.

The weights of male students in a college are modelled by a normal distribution with mean 80 kg and standard deviation 7 kg.

The weights of female students in the college are modelled by a normal distribution with mean 54 kg and standard deviation 5 kg.

The college has a lift installed with a recommended maximum load of 550 kg. One morning, the lift contains 3 male students and 6 female students.

You may assume that the 9 students are randomly chosen.

- a. Find the probability that the weight of a randomly chosen male student is more than twice the weight of a randomly chosen female student. [6]
- b. Determine the probability that their combined weight exceeds the recommended maximum. [5]

## Markscheme

- a. let  $M, F$  denote the weights of the male, female

$$\text{consider } D = M - 2F \quad (\mathbf{M1})$$

$$E(D) = 80 - 2 \times 54 = -28 \quad \mathbf{A1}$$

$$\text{Var}(D) = 7^2 + 4 \times 5^2 \quad (\mathbf{M1})$$

$$= 149 \quad \mathbf{A1}$$

$$P(M > 2F) = P(D > 0) \quad (\mathbf{M1})$$

$$= 0.0109 \quad \mathbf{A1}$$

**Note:** Accept any answer that rounds correctly to 0.011.

**[6 marks]**

- b. consider  $S = \sum_{i=1}^3 M_i + \sum_{i=1}^6 F_i \quad (\mathbf{M1})$

**Note:** Condone the use of the incorrect notation  $3M + 6F$ .

$$E(S) = 3 \times 80 + 6 \times 54 = 564 \quad \mathbf{A1}$$

$$\text{Var}(S) = 3 \times 7^2 + 6 \times 5^2 \quad (\mathbf{M1})$$

$$= 297 \quad \mathbf{A1}$$

$$P(S > 550) = 0.792 \quad \mathbf{A1}$$

**Note:** Accept any answer that rounds correctly to 0.792.

[5 marks]

## Examiners report

- a. [N/A]
- b. [N/A]

Jim is investigating the relationship between height and foot length in teenage boys.

A sample of 13 boys is taken and the height and foot length of each boy are measured.

The results are shown in the table.

<b>Height <math>x</math> cm</b>	129	135	156	146	155	152	139	166	148	179	157	152	160
<b>Foot length <math>y</math> cm</b>	25.8	25.9	29.7	28.6	29.0	29.1	25.3	29.9	26.1	30.0	27.6	27.2	28.0

You may assume that this is a random sample from a bivariate normal distribution.

Jim wishes to determine whether or not there is a positive association between height and foot length.

- a. Calculate the product moment correlation coefficient. [2]
- b. Find the  $p$ -value. [2]
- c. Interpret the  $p$ -value in the context of the question. [1]
- d. Find the equation of the regression line of  $y$  on  $x$ . [2]
- e. Estimate the foot length of a boy of height 170 cm. [2]

## Markscheme

- a. **Note:** In all parts accept answers which round to the correct 2sf answer.

$$r = 0.806 \quad \mathbf{A2}$$

- b.  $4.38 \times 10^{-4} \quad \mathbf{A2}$

- c.  $p$ -value represents strong evidence to indicate a (positive) association between height and foot length  $\mathbf{A1}$

**Note:** FT the  $p$ -value

d.  $y = 0.103x + 12.3$  **A2**

e. attempted substitution of  $x = 170$  **(M1)**

$y = 29.7$  **A1**

**Note:** Accept  $y = 29.8$

## Examiners report

- a. Solutions to this question were often disappointing. Candidates were expected to use appropriate software on their calculators to do the whole question. However, some candidates used their calculators just to evaluate sums and sums of squares and then used the appropriate formulae to calculate the correlation coefficient, the  $p$ -value (which required the evaluation of the  $t$ -value first) and the equation of the regression line. This was a time consuming exercise and introduced the possibility of arithmetic error.
- b. Solutions to this question were often disappointing. Candidates were expected to use appropriate software on their calculators to do the whole question. However, some candidates used their calculators just to evaluate sums and sums of squares and then used the appropriate formulae to calculate the correlation coefficient, the  $p$ -value (which required the evaluation of the  $t$ -value first) and the equation of the regression line. This was a time consuming exercise and introduced the possibility of arithmetic error.
- c. Solutions to this question were often disappointing. Candidates were expected to use appropriate software on their calculators to do the whole question. However, some candidates used their calculators just to evaluate sums and sums of squares and then used the appropriate formulae to calculate the correlation coefficient, the  $p$ -value (which required the evaluation of the  $t$ -value first) and the equation of the regression line. This was a time consuming exercise and introduced the possibility of arithmetic error.
- d. Solutions to this question were often disappointing. Candidates were expected to use appropriate software on their calculators to do the whole question. However, some candidates used their calculators just to evaluate sums and sums of squares and then used the appropriate formulae to calculate the correlation coefficient, the  $p$ -value (which required the evaluation of the  $t$ -value first) and the equation of the regression line. This was a time consuming exercise and introduced the possibility of arithmetic error.
- e. Solutions to this question were often disappointing. Candidates were expected to use appropriate software on their calculators to do the whole question. However, some candidates used their calculators just to evaluate sums and sums of squares and then used the appropriate formulae to calculate the correlation coefficient, the  $p$ -value (which required the evaluation of the  $t$ -value first) and the equation of the regression line. This was a time consuming exercise and introduced the possibility of arithmetic error.

---

Bill is investigating whether or not there is a positive association between the heights and weights of boys of a certain age. He defines the hypotheses

$$H_0 : \rho = 0; H_1 : \rho > 0,$$

where  $\rho$  denotes the population correlation coefficient between heights and weights of boys of this age. He measures the height,  $h$  cm, and weight,  $w$  kg, of each of a random sample of 20 boys of this age and he calculates the following statistics.

$$\sum w = 340, \sum h = 2002, \sum w^2 = 5830, \sum h^2 = 201124, \sum hw = 34150$$

- a. (i) Calculate the correlation coefficient for this sample. [8]
- (ii) Calculate the  $p$ -value of your result and interpret it at the 1% level of significance.
- b. (i) Calculate the equation of the least squares regression line of  $w$  on  $h$ . [3]
- (ii) The height of a randomly selected boy of this age of 90 cm. Estimate his weight.

## Markscheme

a. (i)  $r = \frac{34150 - 340 \times \frac{2002}{20}}{\sqrt{\left(5830 - \frac{340^2}{20}\right)\left(201124 - \frac{2002^2}{20}\right)}} \quad (M1)(A1)$

**Note:** Accept equivalent formula.

$$= 0.610 \quad A1$$

(ii) ( $T = R \times \sqrt{\frac{n-2}{1-R^2}}$  has the  $t$ -distribution with  $n - 2$  degrees of freedom)

$$t = 0.6097666 \dots \sqrt{\frac{18}{1-0.6097666 \dots^2}} \quad M1$$

$$= 3.2640 \dots \quad A1$$

$$DF = 18 \quad A1$$

$$p\text{-value} = 0.00215 \dots \quad A1$$

this is less than 0.01, so we conclude that there is a positive association between heights and weights of boys of this age  $R1$

[8 marks]

- b. (i) the equation of the regression line of  $w$  on  $h$  is

$$w - \frac{340}{20} = \left(\frac{20 \times 34150 - 340 \times 2002}{20 \times 201124 - 2002^2}\right) \left(h - \frac{2002}{20}\right) \quad M1$$

$$w = 0.160h + 0.957 \quad A1$$

(ii) putting  $h = 90$ ,  $w = 15.4$  (kg)  $A1$

**Note:** Award  $M0A0A0$  for calculation of  $h$  on  $w$ .

[3 marks]

## Examiners report

- a. [N/A]  
b. [N/A]

The weights of potatoes in a shop are normally distributed with mean 98 grams and standard deviation 16 grams.

- a. The shopkeeper places 100 randomly chosen potatoes on a weighing machine. Find the probability that their total weight exceeds 10 kilograms. [3]  
kilograms.
- b. Find the minimum number of randomly selected potatoes which are needed to ensure that their total weight exceeds 10 kilograms with probability greater than 0.95. [8]

## Markscheme

- a. let  $T$  denote the total weight, then

$$T \sim N(9800, 25600) \quad (M1)(A1)$$

$$P(T > 10000) = 0.106 \quad A1$$

[3 marks]

- b. let there be  $n$  potatoes, in this case,

$$T \sim N(98n, 256n) \quad A1$$

we require

$$P(T > 10000) > 0.95 \quad (M1)$$

or equivalently

$$P(T \leq 10000) < 0.05 \quad A1$$

standardizing,

$$P\left(Z \leq \frac{10000 - 98n}{16\sqrt{n}}\right) < 0.05 \quad A1$$

$$\frac{10000 - 98n}{16\sqrt{n}} < -1.6449 \dots \quad (A1)$$

$$98n - 26.32\sqrt{n} - 10000 > 0 \quad A1$$

solving the corresponding equation,  $n = 104.7 \dots \quad (A1)$

the required minimum value is 105  $A1$

**Note:** Part (b) could also be solved using SOLVER and normalcdf, or by trial and improvement.

**Note:** Allow the use of = instead of < and > throughout.

[8 marks]

## Examiners report

a. [N/A]

b. [N/A]



The mean weight of a certain breed of bird is claimed to be 5.5 kg. In order to test this claim, a random sample of 10 birds of the breed was obtained and weighed, with the following results in kg.

5.41      5.22      5.54      5.58      5.20      5.57      5.23      5.32      5.46      5.37

You may assume that the weights of this breed of bird are normally distributed.

- a. State suitable hypotheses for testing the above claim using a two-tailed test. [1]
- b. Calculate unbiased estimates of the mean and the variance of the weights of this breed of bird. [4]
- c.i. Determine the  $p$ -value of the above data. [4]
- c.ii. State whether or not the claim is supported by the data, using a significance level of 5%. [1]

## Markscheme

a.  $H_0 : \mu = 5.5; H_1 : \mu \neq 5.5$  **A1**

**[1 mark]**

b.  $\sum x = 53.9, \hat{\mu} = 5.39$  **(M1)A1**

$\sum x^2 = 290.7132, \hat{\sigma}^2 = 0.0214$  **(M1)A1**

**Note:** Accept any answer that rounds correctly to 0.021.

**[4 marks]**

c.i. attempt to use the  $t$ -test **(M1)**

$t = -2.38$  (Accept  $+2.38$ ) **(A1)**

DF = 9 **(A1)**

$p$ -value = 0.0412 **A1**

**[??? marks]**

c.ii. the claim is not supported (not accepted, rejected) at the 5% level of significance **A1**

**[??? marks]**

## Examiners report

- a. [N/A]
  - b. [N/A]
  - c.i. [N/A]
  - c.ii. [N/A]
-

The weights,  $X$  kg, of male birds of a certain species are normally distributed with mean 4.5 kg and standard deviation 0.2 kg. The weights,  $Y$  kg, of female birds of this species are normally distributed with mean 2.5 kg and standard deviation 0.15 kg.

- a. (i) Find the mean and variance of  $2Y - X$ . [6]
- (ii) Find the probability that the weight of a randomly chosen male bird is more than twice the weight of a randomly chosen female bird.
- b. Two randomly chosen male birds and three randomly chosen female birds are placed together on a weighing machine for which the recommended maximum weight is 16 kg. Find the probability that this maximum weight is exceeded. [5]

## Markscheme

- a. (i)  $E(2Y - X) = 2 \times 2.5 - 4.5 = 0.5$  **AI**
- $Var(2Y - X) = 4 \times 0.15^2 + 0.2^2 = 0.13$  **MIAI**
- (ii) We require  $P(X > 2Y) = P(2Y - X < 0)$  **MI**
- 0.0828 **A2**
- Note:** Using tables, answer is 0.0823.

**[6 marks]**

- b. Let  $S$  denote the total weight of the 5 birds.

Then,

$$E(S) = 2 \times 4.5 + 3 \times 2.5 = 16.5 \quad \mathbf{AI}$$
$$Var(S) = 2 \times 0.2^2 + 3 \times 0.15^2 = 0.1475 \quad \mathbf{MIAI}$$
$$P(S > 16) = 0.904 \quad \mathbf{A2}$$

**Note:** Using tables, answer is 0.903.

**[5 marks]**

## Examiners report

- a. [N/A]  
b. [N/A]

---

Sarah is the quality control manager for the Stronger Steel Corporation which makes steel sheets. The steel sheets should have a mean tensile strength of 430 MegaPascals (MPa). If the mean tensile strength drops to 400 MPa, then Sarah must recommend a change in composition. The tensile strength of these steel sheets follows a normal distribution with a standard deviation of 35 MPa. Sarah defines the following hypotheses

$$H_0 : \mu = 430$$

$$H_1 : \mu = 400$$

where  $\mu$  denotes the mean tensile strength in MPa. She takes a random sample of  $n$  steel sheets and defines the critical region as  $\bar{x} \leq k$ , where  $\bar{x}$  notes the mean tensile strength of the sample in MPa and  $k$  is a constant.

Given that the  $P(\text{Type I Error}) = 0.0851$  and  $P(\text{Type II Error}) = 0.115$ , both correct to three significant figures, find the value of  $k$  and the value of  $n$ .

## Markscheme

$$\bar{X} \sim N\left(430, \frac{35^2}{n}\right) \quad (M1)(A1)$$

**Note:** The M1 is for considering the distribution of  $\bar{X}$

type I error gives  $P(\bar{X} \leq k/\mu = 430) = 0.0851$

$$\frac{k-430}{\frac{35}{\sqrt{n}}} = -1.37156 \dots \quad M1A1$$

type II error gives  $P(\bar{X} > k/\mu = 400) = 0.115$

$$\frac{k-400}{\frac{35}{\sqrt{n}}} = 1.20035 \dots \quad M1A1$$

**Note:** The two M1 marks above are for attempting to standardize  $\bar{X}$  and obtain the corresponding equations with inverse normal values

solving simultaneously (M1)

$$k = 414 \quad A1$$

$$n = 9 \quad A1$$

## Examiners report

This proved to be a difficult question for most candidates with only a minority giving a correct solution. Most candidates either made no attempt at the question or just wrote several lines of irrelevant mathematics.

a. Bottles of iced tea are supposed to contain 500 ml. A random sample of 8 bottles was selected and the volumes measured (in ml) were as [5]

follows:

497.2, 502.0, 501.0, 498.6, 496.3, 499.1, 500.1, 497.7 .

(i) Calculate unbiased estimates of the mean and variance.

(ii) Test at the 5% significance level the null hypothesis  $H_0 : \mu = 500$  against the alternative hypothesis  $H_1 : \mu < 500$  .

b. A random sample of size four is taken from the distribution  $N(60, 36)$  . [6]

Calculate the probability that the sum of the sample values is less than 250.

## Markscheme

a. (i) 497.2, 502.0, 501.0, 498.6, 496.3, 499.1, 500.1, 497.7

using the GDC

$$\bar{x} = 499.0, \sigma^2 = 3.8(0) \quad \mathbf{AIAI}$$

**Note:** Accept 499.

(ii) **EITHER**

$$p\text{-value} = 0.0950 \quad \mathbf{AI}$$

since  $0.0950 > 0.05$  accept  $H_0$   $\mathbf{RIAI}$

**OR**

$$t_{calc} = -1.45, t_{critical} = -1.895 \text{ for } v = 7 \text{ at } 5\% \text{ level} \quad \mathbf{AI}$$

since  $t_{calc} > t_{critical}$  accept  $H_0$   $\mathbf{RIAI}$

**[5 marks]**

b. each  $X \sim N(60, 36)$  so  $\sum_{n=1}^4 X_n \sim N(4(60), 4(36)) = N(240, 144)$   $\mathbf{MIAIAI}$

$$\Pr(\text{Sum} < 250) = \Pr\left(z < \frac{250-240}{\sqrt{144}} = \frac{5}{6}\right) \quad \mathbf{(MI)(AI)}$$

$$= 0.798 \text{ (by GDC)} \quad \mathbf{AI}$$

**Notes:** Accept 0.797 (tables).

Answer only is awarded  $\mathbf{M0A0A0(MI)(AI)AI}$ .

**[6 marks]**

## Examiners report

- a. (a)(i) Very few mistakes were made in this question, although sometimes variance and standard deviation were confused. Why both variance and standard deviation are needed might be something that teachers could explore.
- (ii) Again there were no serious problems although some candidates fail to show all the important parameters such as degrees of freedom.
- b. This was found to be relatively straightforward except for using the correct variance of 144. It would be useful here to make clear the distinction between the sum of random variables and a multiple of a random variable.

---

The discrete random variables  $X_n, n \in \mathbb{Z}^+$  have probability generating functions given by  $G_n(t) = \frac{t}{n} \left( \frac{t^n - 1}{t - 1} \right)$ .

Let  $X_{n-1}$  and  $X_{n+1}$  be independent.

a. Use the formula for the sum of a finite geometric series to show that

[4]

$$P(X_n = k) = \begin{cases} \frac{1}{n} & \text{for } 1 \leq k \leq n \\ 0 & \text{otherwise} \end{cases}.$$

b. Find  $E(X_n)$ .

[3]

c. Find the set of values of  $n$  for which  $E(X_{n-1} \times X_{n+1}) < 2n$ .

[4]

# Markscheme

a. using  $\left(\frac{t^n-1}{t-1}\right) = 1 + t + t^2 + \dots + t^{n-1}$  **M1**

$$G_n(t) = 0 + \frac{t}{n} + \frac{t^2}{n} + \frac{t^3}{n} + \dots + \frac{t^n}{n} + 0 \times t^{n+1} + 0 \times \dots \quad \mathbf{A1A1}$$

**Note:** **A1** for the non-zero terms, **A1** for the observation that all other terms are zero.

the statement that the coefficient of  $t^k$  gives  $P(X_n = k)$  **R1**

$$\text{hence } P(X_n = k) = \begin{cases} \frac{1}{n} & \text{for } 1 \leq k \leq n \\ 0 & \text{otherwise} \end{cases} \quad \mathbf{AG}$$

**[4 marks]**

b.  $E(X_n) = 0 \times 0 + 1 \times \frac{1}{n} + 2 \times \frac{1}{n} + 3 \times \frac{1}{n} + \dots + n \times \frac{1}{n} + (n+1) \times 0 + \dots \times 0$  **(M1)(A1)**

$$= \frac{1}{n} \times \sum_{k=1}^{k=n} k$$

$$= \frac{1}{n} \times \frac{1}{2}n(n+1) = \frac{n+1}{2} \quad \mathbf{A1}$$

**Note:** Accept use of  $G'(1)$ .

**[3 marks]**

c.  $X_{n-1}$  and  $X_{n+1}$  are independent  $\Rightarrow E(X_{n-1} \times X_{n+1}) = E(X_{n-1}) \times E(X_{n+1})$  **M1**

$$= \frac{n}{2} \times \frac{n+2}{2} \quad \mathbf{A1}$$

required to solve  $n^2 < 6n$  (or  $n + 2 < 8$ ) **M1**

solution:  $(2 \leq) n < 6$  **A1**

**[4 marks]**

# Examiners report

a. Again this was a question that tested candidates and although many started only a very limited number were able to make significant progress.

Part (a) was rarely done well with most candidates not understanding what was required. There was a little more success with part (b) but a number of candidates attempted methods that were not going to lead to anything meaningful. Most candidates did not understand what was required from part (c) and few correct answers were seen, even taking into account the fact that follow through marks could be awarded from (b).

b. Again this was a question that tested candidates and although many started only a very limited number were able to make significant progress.

Part (a) was rarely done well with most candidates not understanding what was required. There was a little more success with part (b) but a number of candidates attempted methods that were not going to lead to anything meaningful. Most candidates did not understand what was required from part (c) and few correct answers were seen, even taking into account the fact that follow through marks could be awarded from (b).

c. Again this was a question that tested candidates and although many started only a very limited number were able to make significant progress.

Part (a) was rarely done well with most candidates not understanding what was required. There was a little more success with part (b) but a number of candidates attempted methods that were not going to lead to anything meaningful. Most candidates did not understand what was required from part (c) and few correct answers were seen, even taking into account the fact that follow through marks could be awarded from (b).

---

Bill buys two biased coins from a toy shop.

a. The shopkeeper claims that when one of the coins is tossed, the probability of obtaining a head is 0.6. Bill wishes to test this claim by [6]  
tossing the coin 250 times and counting the number of heads obtained.

(i) State suitable hypotheses for this test.

(ii) He obtains 140 heads. Find the  $p$ -value of this result and determine whether or not it supports the shopkeeper's claim at the 5% level of significance.

b. Bill tosses the other coin a large number of times and counts the number of heads obtained. He correctly calculates a 95% confidence [7]  
interval for the probability that when this coin is tossed, a head is obtained. This is calculated as [0.35199, 0.44801] where the end-points are correct to five significant figures.

Determine

(i) the number of times the coin was tossed;

(ii) the number of heads obtained.

## Markscheme

a. (i)  $H_0 : p = 0.6 ; H_1 : p \neq 0.6$  *A1A1*

(ii) **EITHER**

using a normal approximation,  $p$ -value = 0.197 *A2*

**Note:** Award *AI* for 0.0984.

the shopkeeper's claim is supported *AI*

because  $0.197 > 0.05$  *RI*

**OR**

using binomial distribution,  $p$ -value = 0.221 *A2*

**Note:** Award *AI* for 0.110.

the shopkeeper's claim is supported *AI*

because  $0.221 > 0.05$  *RI*

**Note:** Follow through the candidate's  $p$ -value for *AIRI*.

**Note:** Accept  $p$ -values correct to two significant figures.

[6 marks]

b. (i)  $\hat{p} = \frac{0.35199+0.44801}{2} = 0.4$  *AI*

width of CI =  $3.92\sqrt{\frac{0.4 \times 0.6}{n}}$  *MI*

$3.92\sqrt{\frac{0.4 \times 0.6}{n}} = 0.44801 - 0.35199 = 0.096(02)$  *AI*

solving,

$n = \left(\frac{3.92}{0.096(02)}\right)^2 \times 0.24$  *(MI)*

= 400 *AI*

(ii)  $x = n\hat{p} = 400 \times 0.4 = 160$  *MI AI*

[7 marks]

## Examiners report

- a. Part (a) was well answered in general, using the calculator either to carry out a significance test on proportions or to find the  $p$ -value directly using the binomial distribution. Some candidates gave their conclusion in the form "Accept  $H_0$ ", this was not accepted since the question asked whether or not the shopkeeper's claim was supported and a direct answer to this question was required.
- b. Part (b) caused problems for some candidates who were unsure how to proceed. Some candidates used a trial and error method which involved showing that  $\hat{p} = 0.4$  and then using their calculator to find the confidence interval for appropriate pairs of values for  $n$  and  $p$  until reaching 400, 160. This was accepted as a valid method although it is not recommended as a general method since its success was based upon the value of  $n$  being one that would probably be tested.

---

All members of a large athletics club take part in an annual shotput competition.

The following data give the distances achieved, in metres, by a random selection of 10 members of the club in the 2016 competition

11.8, 14.3, 13.8, 10.3, 14.9, 14.7, 12.4, 13.9, 14.0, 11.7

The president of the club wishes to test whether these data provide evidence that distances achieved have increased since the 2015 competition, when the mean result for the club was 12.4 m. You may assume that the distances achieved follow a normal distribution with mean  $\mu$ , variance  $\sigma^2$ , and that the membership of the club has not changed from 2015 to 2016.

- a. State suitable hypotheses. [1]
- b. (i) Give a reason why a  $t$  test is appropriate and write down its degrees of freedom. [4]
- (ii) Find the critical region for testing at each of the 5% and 10% significance levels.
- c. (i) Find unbiased estimates of  $\mu$  and  $\sigma^2$ . [3]
- (ii) Find the value of the test statistic.

d. State the conclusions that the president of the club should reach from this test, giving reasons for your answer.

[2]

## Markscheme

a.  $H_0 : \mu = 12.4$ ;  $H_1 : \mu > 12.4$  **A1**

**[1 mark]**

b. (i)  $t$  test is appropriate because the variance (standard deviation) is unknown **R1**

$v = 9$  **A1**

(ii)  $t \geq 1.83$  (5%);  $t \geq 1.38$  (10%) **A1A1**

**Note:** Accept strict inequalities.

**[4 marks]**

c. (i) unbiased estimate of  $\mu$  is 13.18 **A1**

**Note:** Accept 13.2.

unbiased estimate of  $\sigma^2$  is 2.34 ( $1.531^2$ ) **A1**

(ii)  $t_{\text{calc}} = \left( \frac{13.18 - 12.4}{\frac{1.531}{\sqrt{10}}} \right) = 1.61$  or 1.65 **A1**

**[3 marks]**

d. as  $1.38 < 1.61 < 1.83$  **R1**

evidence to accept  $H_0$  at the 5% level, but not at the 10% level **A1**

**Note:** Accept the use of the  $p$ -value = 0.0708.

**[2 marks]**

## Examiners report

a. Most candidates had an understanding of how to start the question, but only a small number were able to gain full marks. It appeared that many candidates were used to finding  $p$ -values, but showed a lack of understanding when asked to find the critical regions and test a  $t$ -value. The conclusions required in part (d) were often too brief and/or poorly expressed.

b. Most candidates had an understanding of how to start the question, but only a small number were able to gain full marks. It appeared that many candidates were used to finding  $p$ -values, but showed a lack of understanding when asked to find the critical regions and test a  $t$ -value. The conclusions required in part (d) were often too brief and/or poorly expressed.



- c. Most candidates had an understanding of how to start the question, but only a small number were able to gain full marks. It appeared that many candidates were used to finding  $p$ -values, but showed a lack of understanding when asked to find the critical regions and test a  $t$ -value. The conclusions required in part (d) were often too brief and/or poorly expressed.
- d. Most candidates had an understanding of how to start the question, but only a small number were able to gain full marks. It appeared that many candidates were used to finding  $p$ -values, but showed a lack of understanding when asked to find the critical regions and test a  $t$ -value. The conclusions required in part (d) were often too brief and/or poorly expressed.

- a. Sami is undertaking market research on packets of soap powder. He considers the brand “Gleam”. The weight of the contents of a randomly chosen packet of “Gleam” follows a normal distribution with mean 750 grams and standard deviation 20 grams. [8]
- The weight of the packaging follows a different normal distribution with mean 40 grams and standard deviation 5 grams.
- Find:
- the probability that a randomly chosen packet of “Gleam” has a **total** weight exceeding 780 grams.
  - the probability that the total weight of the **contents** of five randomly chosen packets of “Gleam” exceeds 3800 grams.
- b. Sami now considers the brand “Bright”. The weight of the contents of a randomly chosen packet of “Bright” follow a normal distribution with mean 650 grams and standard deviation 16 grams. Find the probability that the **contents** of six randomly chosen packets of “Bright” weigh more than the **contents** of five randomly chosen packets of “Gleam”. [4]

## Markscheme

- a. **Note:** In all parts accept answers which round to the correct 2sf answer.

(i) contents:  $X \sim N(750, 400)$

packaging:  $Y \sim N(40, 25)$

consider  $X + Y$  **(M1)**

$$E(X + Y) = 790 \quad \mathbf{A1}$$

$$\text{Var}(X + Y) = 425 \quad \mathbf{A1}$$

$$P(X + Y > 780) = 0.686 \quad \mathbf{A1}$$

(ii) Let  $X_1 + X_2 + X_3 + X_4 + X_5 = A$  **M1**

$$E(A) = 5E(X) = 3750 \quad \mathbf{A1}$$

$$\text{Var}(A) = 5\text{Var}(X) = 2000 \quad \mathbf{A1}$$

$$P(A > 3800) = 0.132 \quad \mathbf{A1}$$

**Note:** Condone the notation  $A = 5X$  if the variance is correct, **MO** if not

- b. contents of Bright:  $B \sim N(650, 256)$

$$\text{let } G = B_1 + B_2 + B_3 + B_4 + B_5 + B_6 - (X_1 + X_2 + X_3 + X_4 + X_5) \quad \mathbf{M1}$$

$$E(G) = 6 \times 650 - 5 \times 750 = 150 \quad \mathbf{A1}$$

$$\text{Var}(G) = 6 \times 256 + 5 \times 400 = 3536 \quad \mathbf{A1}$$

$$P(G > 0) = 0.994 \quad \mathbf{A1}$$

**Note:** Condone the notation  $G = 6B - 5X$  if the variance is correct, **MO** if not

## Examiners report

- a. Part (a)(i) was well answered in general. In (a)(ii) and (b), however, many candidates made the fairly common error of confusing  $\sum_{i=1}^n X_i$  with  $nX$  which gives an incorrect variance. This is an important distinction which needs to be emphasized.
- b. Part (a)(i) was well answered in general. In (a)(ii) and (b), however, many candidates made the fairly common error of confusing  $\sum_{i=1}^n X_i$  with  $nX$  which gives an incorrect variance. This is an important distinction which needs to be emphasized.

The random variables  $X, Y$  follow a bivariate normal distribution with product moment correlation coefficient  $\rho$ . The following table gives a random sample from this distribution.

$x$	5.1	3.8	3.7	2.5	4.0	3.7	1.6	2.8	3.3	2.9
$y$	4.6	4.9	4.1	5.9	4.2	1.6	5.1	2.1	6.4	4.7

- (a) Determine the value of  $r$ , the product moment correlation coefficient of this sample.
- (b) (i) Write down hypotheses in terms of  $\rho$  which would enable you to test whether or not  $X$  and  $Y$  are independent.  
(ii) Determine the  $p$ -value of the above sample and state your conclusion at the 5% significance level. Justify your answer.
- (c) (i) Determine the equation of the regression line of  $y$  on  $x$ .  
(ii) State whether or not this equation can be used to obtain an accurate prediction of the value of  $y$  for a given value of  $x$ . Give a reason for your answer.

## Markscheme

(a)  $r = -0.163 \quad \mathbf{A2}$

[2 marks]

(b) (i)  $H_0 : \rho = 0 : H_1 : \rho \neq 0 \quad \mathbf{A1}$

(ii)  $t = r \sqrt{\frac{n-2}{1-r^2}} = -0.468 \dots \quad \mathbf{(A1)}$

DF = 8  $\quad \mathbf{(A1)}$

$p$ -value =  $2 \times 0.326 \dots = 0.652 \quad \mathbf{A1}$

since  $0.652 > 0.05$ , we accept  $H_0 \quad \mathbf{R1}$

**Note:** Award  $\mathbf{(A1)(A1)A0}$  if the  $p$ -value is given as 0.326 without prior working.

**Note:** Follow through their  $p$ -value for the  $\mathbf{R1}$ .

[5 marks]

(c) (i)  $y = -0.257x + 5.22$  **AI**

**Note:** Accept answers which round to  $-0.26$  and  $5.2$ .

(ii) no, because  $X$  and  $Y$  have been shown to be independent (or equivalent) **AI**

[2 marks]

## Examiners report

[N/A]

The following table shows the probability distribution of the discrete random variable  $X$ .

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

(a) Show that the probability generating function of  $X$  is given by

$$G(t) = \frac{t(1+t)^2}{4}.$$

(b) Given that  $Y = X_1 + X_2 + X_3 + X_4$ , where  $X_1, X_2, X_3, X_4$  is a random sample from the distribution of  $X$ ,

(i) state the probability generating function of  $Y$ ;

(ii) hence find the value of  $P(Y = 8)$ .

## Markscheme

(a)  $G(t) = \frac{1}{4}t + \frac{1}{2}t^2 + \frac{1}{4}t^3$  **MIAI**

$$= \frac{t(1+t)^2}{4} \quad \mathbf{AG}$$

[2 marks]

(b) (i) PGF of  $Y = (G(t))^4 = \left(\frac{t(1+t)^2}{4}\right)^4$  **AI**

(ii)  $P(Y = 8) = \text{coefficient of } t^8$  **(MI)**

$$= \frac{{}^8C_4}{256} \quad \mathbf{(AI)}$$

$$= \frac{35}{128}(0.273) \quad \mathbf{AI}$$

**Note:** Accept 0.27 or answers that round to 0.273.

[4 marks]

## Examiners report

Let  $X_k$  be independent normal random variables, where  $E(X_k) = \mu$  and  $Var(X_k) = \sqrt{k}$ , for  $k = 1, 2, \dots$ .

The random variable  $Y$  is defined by  $Y = \sum_{k=1}^6 \frac{(-1)^{k+1}}{\sqrt{k}} X_k$ .

- a. (i) Find  $E(Y)$  in the form  $p\mu$ , where  $p \in \mathbb{R}$ . [5]
- (ii) Find  $k$  if  $Var(X_k) < Var(Y) < Var(X_{k+1})$ .
- b. A random sample of  $n$  values of  $Y$  was found to have a mean of 8.76. [6]
- (i) Given that  $n = 10$ , determine a 95% confidence interval for  $\mu$ .
- (ii) The width of the confidence interval needs to be halved. Find the appropriate value of  $n$ .

## Markscheme

a. (i)  $E(Y) = \frac{1}{\sqrt{1}}\mu - \frac{1}{\sqrt{2}}\mu + \frac{1}{\sqrt{3}}\mu - \frac{1}{\sqrt{4}}\mu + \frac{1}{\sqrt{5}}\mu - \frac{1}{\sqrt{6}}\mu$  (M1)

$= 0.409 (209)\mu$  AI

**Note:** Accept answers which round to 0.41.

(ii)  $Var(Y) = \frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{4}} + \frac{1}{\sqrt{5}} + \frac{1}{\sqrt{6}}$  (M1)

$= 3.64 (3.6399\dots)$  AI

$(Var(X_{13}) = 3.61; Var(X_{14}) = 3.74) \Rightarrow k = 13$  AI

[5 marks]

- b. (i) 95% CI for  $E(Y)$

$8.76 \pm 1.96 \sqrt{\frac{3.6399\dots}{10}}$  (M1) □□

$= [7.58, 9.94]$  AIAI

**Note:** Accept [7.6, 9.9]. Do not penalize answers given to more than 3sf.

Since  $\mu = \frac{E(Y)}{0.409\dots}$ , CI for  $\mu$  is [18.5, 24.3] AI

**Note:** Do not penalize answers given to more than 3sf.

- (ii) width of a CI is inversely proportional to the square root of  $n$  (M1)

so  $n = 40$  AI

[6 marks]

# Examiners report

- a. This question proved to be the most difficult on the paper and few fully correct answers were seen. In part (a) (i) many candidates did know how to find the answers in terms of  $\mu$ . Very few candidates successfully completed part (a) (ii).
- b. This question proved to be the most difficult on the paper and few fully correct answers were seen. In part b) (i) a number of candidates made some progress, but few realised or knew how to convert the confidence interval for  $E(Y)$  into a confidence interval for  $\mu$ . For those who persevered to the end of the question, there was a reasonable degree of success in part (b) (ii).

The weights, in grams, of 10 apples were measured with the following results:

212.2 216.9 209.0 215.5 215.9 213.5 208.9 213.8 216.4 209.9

You may assume that this is a random sample from a normal distribution with mean  $\mu$  and variance  $\sigma^2$ .

- (a) Giving all your answers correct to four significant figures,  
(i) determine unbiased estimates for  $\mu$  and  $\sigma^2$ ;  
(ii) find a 95% confidence interval for  $\mu$ .

Another confidence interval for  $\mu$ , [211.5, 214.9], was calculated using the above data.

- (b) Find the confidence level of this interval.

## Markscheme

(a) (i)  $\bar{x} = 213.2$  *AI*

$s = 3.0728\dots$  *(AI)*

$s^2 = 9.442$  *AI*

(ii) [211.0, 215.4] *AI AI*

**Note:** Accept 211 in place of 211.0.

**Note:** Apart from the above note, accept any answers which round to the correct 4 significant figure answers.

*[5 marks]*

(b) use of the fact that the width of the interval is  $2t \times \frac{s}{\sqrt{n}}$  *(AI)*

so that  $3.4 = 2t \times \frac{3.0728\dots}{\sqrt{10}}$  *MI*

$t = 1.749$  *AI*

degrees of freedom = 9 *(AI)*

$P(T > 1.749) = 0.0571$  *(MI)*

confidence level =  $1 - 2 \times 0.0571 = 0.886$  (88.6%) *AI*

**Note:** Award the DF = 9 *(AI)* mark if the following line has 0.00337 on the RHS.

**Note:** Accept any answer which rounds to 88.6%.

[6 marks]

# Examiners report

[N/A]

A sample of size 100 is taken from a normal population with unknown mean  $\mu$  and known variance 36.

Another investigator decides to use the same data to test the hypotheses  $H_0: \mu = 65$ ,  $H_1: \mu = 67.9$ .

a. An investigator wishes to test the hypotheses  $H_0: \mu = 65$ ,  $H_1: \mu > 65$ . [3]

He decides on the following acceptance criteria:

Accept  $H_0$  if the sample mean  $\bar{x} \leq 66.5$

Accept  $H_1$  if  $\bar{x} > 66.5$

Find the probability of a Type I error.

b.i. She decides to use the same acceptance criteria as the previous investigator. Find the probability of a Type II error. [3]

b.ii. Find the critical value for  $\bar{x}$  if she wants the probabilities of a Type I error and a Type II error to be equal. [3]

## Markscheme

a.  $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$$\bar{X} \sim N\left(65, \frac{36}{100}\right) \quad (\mathbf{A1})$$

$$P(\text{Type I Error}) = P(\bar{X} > 66.5) \quad (\mathbf{M1})$$

$$= 0.00621 \quad \mathbf{A1}$$

[3 marks]

b.i.  $P(\text{Type II Error}) = P(\text{accept } H_0 | H_1 \text{ is true})$

$$= P(\bar{X} \leq 66.5 | \mu = 67.9) \quad (\mathbf{M1})$$

$$= P(\bar{X} \leq 66.5) \text{ when } \bar{X} \sim N\left(67.9, \frac{36}{100}\right) \quad (\mathbf{M1})$$

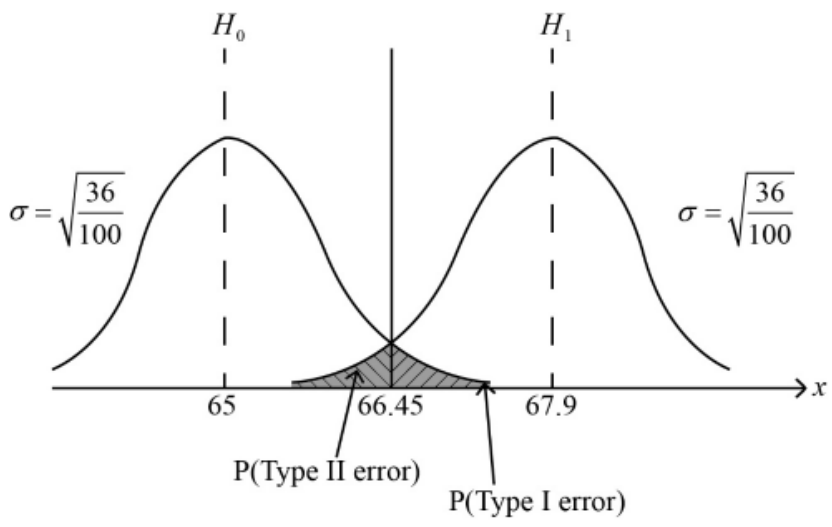
$$= 0.00982 \quad \mathbf{A1}$$

[3 marks]

b.ii. the variances of the distributions given by  $H_0$  and  $H_1$  are equal, (R1)

by symmetry the value of  $\bar{x}$  lies midway between 65 and 67.9 (M1)

$$\Rightarrow \bar{x} = \frac{1}{2}(65 + 67.9) = 66.45 \quad \mathbf{A1}$$



[3 marks]

## Examiners report

- a. [N/A]
- b.i. [N/A]
- b.ii. [N/A]