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 μ for this paper. [4]

b. The null hypothesis $\mu = 46.5$ is tested against the alternative hypothesis $\mu < 46.5$ at the λ% significance level. Determine the set of values of λ for which the null hypothesis is rejected in favour of the alternative hypothesis. [4]

A random sample $X_1,\ X_2,\ \ldots,\ X_n$ is taken from the normal distribution $\mathrm{N}(\mu,\ \sigma^2)$, where the value of μ is unknown but the value of σ^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 μ,

$$[3.12,\ 3.25].$$

The teacher asks Alun to interpret this result. Alun makes the following statement. “The value of μ 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]

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

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

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

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

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

a. Find $\mathrm{P}\left(B > \mathrm{E}(B) + \frac{1}{2}\sqrt{\mathrm{Var}(B)}\right)$. [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]

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]

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.

Height x cm	129	135	156	146	155	152	139	166	148	179	157	152	160
Foot length y cm	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]

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

$$\mathrm{H}_0 : \rho = 0; \mathrm{H}_1 : \rho > 0,$$

where ρ 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.

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]

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]

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]

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]

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 μ denotes the mean tensile strength in MPa. She takes a random sample of n steel sheets and defines the critical region as $\bar{x} \leqslant k$, 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.

---

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 follows: [5]

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

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

(ii) Test at the 5% significance level the null hypothesis $\mathrm{H}_0 : \mu = 500$ against the alternative hypothesis $\mathrm{H}_1 : \mu < 500$.

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.

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]

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

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]

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 tossing the coin 250 times and counting the number of heads obtained. [6]

(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 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. [7]

Determine

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

(ii) the number of heads obtained.

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 μ, variance σ^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 μ and σ^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]

---

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:

(i) the probability that a randomly chosen packet of “Gleam” has a **total** weight exceeding 780 grams.

(ii) 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]

The random variables X, Y follow a bivariate normal distribution with product moment correlation coefficient ρ. 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 ρ 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.

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

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

(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 $\mathrm{P}(Y = 8)$.

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

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

a. (i) Find $\mathrm{E}(Y)$ in the form $p\mu$, where $p \in \mathbb{R}$. [5]

(ii) Find k if $\mathrm{Var}(X_k) < \mathrm{Var}(Y) < \mathrm{Var}(X_{k+1})$.

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 μ .

(ii) The width of the confidence interval needs to be halved. Find the appropriate value of n .

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 μ and variance σ^2.

(a) Giving all your answers correct to four significant figures,

(i) determine unbiased estimates for μ and σ^2;

(ii) find a 95% confidence interval for μ.

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

(b) Find the confidence level of this interval.

A sample of size 100 is taken from a normal population with unknown mean μ 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]