Dr Oliver Mathematics
Advanced Subsidiary Paper 21: Statistics
June 2022: Calculator
1 hour 15 minutes

The total number of marks available is 30.
You must write down all the stages in your working.
Inexact answers should be given to three significant figures unless otherwise stated.
(It goes with Paper 22: Mechanics)

1. The relationship between two variables p and t is modelled by the regression line with equation
$$p = 22 - 1.1t.$$

   The model is based on observations of the independent variable, t, between 1 and 10.

   (a) Describe the correlation between p and t implied by this model. (1)

   > **Solution**
   > E.g., <u>negative correlation</u>.

   Given that p is measured in centimetres and t is measured in days,

   (b) state the units of the gradient of the regression line. (1)

   > **Solution**
   > E.g., <u>cm/day</u>.

   Using the model,

   (c) calculate the change in p over a 3-day period. (2)

   > **Solution**
   > $$p = 22 - 1.1t \Rightarrow \frac{\mathrm{d}p}{\mathrm{d}t} = -1.1$$
   > and
   > $$t = 3 \Rightarrow \frac{\mathrm{d}p}{\mathrm{d}t} = -3.3;$$
   > hence, it is a <u>decrease</u> of <u>3.3 cm</u>.

   Tisam uses this model to estimate the value of p when $t = 19$.

(d) Comment, giving a reason, on the reliability of this estimate. (1)

> **Solution**
> E.g., 19 is well outside the range $1 \leqslant t \leqslant 10$, it involves extrapolation so it is unreliable/inaccurate, etc

2. A manufacturer of sweets knows that 8% of the bags of sugar delivered from supplier A will be damp.

A random sample of 35 bags of sugar is taken from supplier A.

(a) Using a suitable model, find the probability that the number of bags of sugar that are damp is (3)

   (i) exactly 2,

> **Solution**
> Let D be the number of bags that are damp: $D \sim \mathrm{B}(35, 0.08)$. Now,
> $$P(D = 2) = \binom{35}{2}(0.08)^2(0.92)^{33}$$
> $$= 0.243\,049\,855 \text{ (FCD)}$$
> $$= \underline{\underline{0.243}} \text{ (3 sf)}.$$

   (ii) more than 3.

> **Solution**
> $$P(D > 3) = 1 - P(D \leqslant 3)$$
> $$= 1 - 0.693\,970\,630\,7 \text{ (FCD)}$$
> $$= 0.306\,029\,363\,3 \text{ (FCD)}$$
> $$= \underline{\underline{0.306}} \text{ (3 sf)}.$$

Supplier B claims that when it supplies bags of sugar, the proportion of bags that are damp is less than 8%.

The manufacturer takes a random sample of 70 bags of sugar from supplier B and finds that only 2 of the bags are damp.

(b) Carry out a suitable test to assess supplier B's claim. (4)
   You should state your hypotheses clearly and use a 10% level of significance.

Solution

Let E be the number of bags that are damp: $E \sim \mathrm{B}(70, 0.08)$.
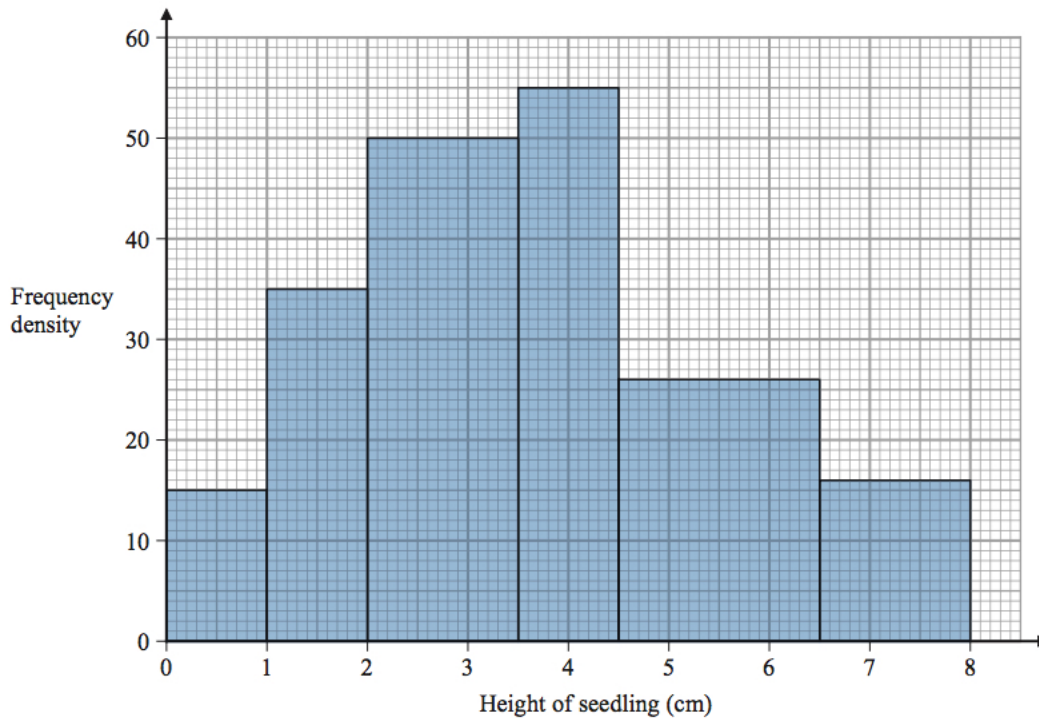
H_0: $p = 0.08$.

H_1: $p < 0.08$.

Level of significance: 0.01

Now,

$$\mathrm{P}(E \leqslant 2) = 0.073\,975\,623\,19 < 0.10;$$

Now, the result is significant so we reject H_0 and we conclude that there is evidence to <u>support</u> supplier B's claim.

3. The histogram summarises the heights of 256 seedlings two weeks after they were planted.

Height of seedling (cm)

(a) Use linear interpolation to estimate the median height of the seedlings. (4)

Solution

We make up a table:

3

Interval	Frequency	Cumulative Frequency
$0-1$	$15 \times 1 = 15$	15
$1-2$	$35 \times 1 = 35$	$15 + 35 = 50$
$2-3.5$	$50 \times 1.5 = 75$	$75 + 50 = 125$
$3.5-4.5$	$55 \times 1 = 55$	$55 + 125 = 180$
$4.5-6.5$	$26 \times 2 = 52$	$52 + 180 = 232$
$6.5-8$	$16 \times 1.5 = 24$	$24 + 232 = 256$

Now, the median is at
$$\tfrac{1}{2}(256 + 1) = 128.5\text{th}$$
which is 3.5th member of the $3.5 - 4.5$ group:

$$\text{median} = 3.5 + \frac{3.5}{55} \times 1$$
$$= 3\tfrac{31}{55} \text{ (exact)}$$
$$= \underline{\underline{3.56 \text{ (3 sf)}}}.$$

Chris decides to model the **frequency density** for these 256 seedlings by a curve with equation
$$y = kx(8 - x),$$
where k is a constant.

(b) Find the value of k. (3)

Solution

$$\int_0^8 kx(8 - x)\,\mathrm{d}x = 256 \Rightarrow k\int_0^8 (8x - x^2)\,\mathrm{d}x = 256$$
$$\Rightarrow k\left[4x^2 - \tfrac{1}{3}x^3\right]_{x=0}^8 = 256$$
$$\Rightarrow k\left\{(256 - 170\tfrac{2}{3}) - (0 - 0)\right\} = 256$$
$$\Rightarrow \tfrac{256}{3}k = 256$$
$$\Rightarrow \underline{\underline{k = 3}}.$$

Using this model,

(c) write down the median height of the seedlings. (1)

> **Solution**
>
> 4.

4. Jiang is studying the variable Daily Mean Pressure from the large data set.
   (You are not expected to have memorised values from the large data set. The question is simply looking for sensible answers.)

   He drew the following box and whisker plot for these data for one of the months for one location using a linear scale but

   - he failed to label all the values on the scale and
   - he gave an incorrect value for the median.

   1200

   Daily Mean Pressure (hPa)
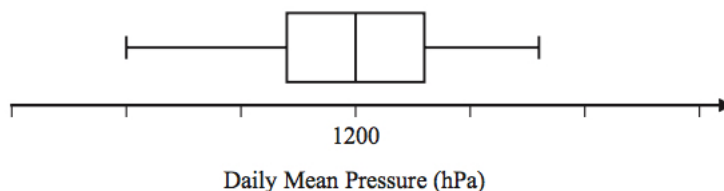
   Using your knowledge of the large data set, suggest a suitable value for

   (a) the median, (1)

   > **Solution**
   > E.g., 990 to 1030 inclusive.

   (b) the range. (1)

   > **Solution**
   > E.g., any range between 10 and 50 inclusive.

5. Manon has two biased spinners, one red and one green.

   The random variable R represents the score when the red spinner is spun.
   The random variable G represents the score when the green spinner is spun.

   The probability distributions for R and G are given below.

r	2	3
$P(R = r)$	$\frac{1}{4}$	$\frac{3}{4}$

g	1	4
$\mathrm{P}(G = g)$	$\frac{2}{3}$	$\frac{1}{3}$

Manon spins each spinner once and adds the two scores.

(a) Find the probability that $\hspace{8cm}$ (3)

    (i) the sum of the two scores is 7,

> **Solution**
> $$\mathrm{P}(R = 3, G = 4) = \mathrm{P}(R = 3) \times \mathrm{P}(G = 4)$$
> $$= \tfrac{3}{4} \times \tfrac{1}{3}$$
> $$= \underline{\underline{\tfrac{1}{4}}}.$$

    (ii) the sum of the two scores is less than 4.

> **Solution**
> $$\mathrm{P}(R = 2, G = 1) = \mathrm{P}(R = 2) \times \mathrm{P}(G = 1)$$
> $$= \tfrac{1}{4} \times \tfrac{2}{3}$$
> $$= \underline{\underline{\tfrac{1}{6}}}.$$

The random variable
$$X = mR + nG,$$
where m and n are integers.

$$\mathrm{P}(X = 20) = \tfrac{1}{6} \text{ and } \mathrm{P}(X = 50) = \tfrac{1}{4}.$$

(b) Find the value of m and the value of n. $\hspace{6cm}$ (5)

> **Solution**
> Well,
> $$\mathrm{P}(X = 20) = \tfrac{1}{6} \Rightarrow R = 2, \, G = 1$$
> $$\Rightarrow 2m + n = 20 \quad (1)$$

and

$$P(X = 50) = \tfrac{1}{4} \Rightarrow R = 3,\ G = 4$$
$$\Rightarrow 3m + 4n = 50 \quad (2).$$

Do $4 \times (1)$:

$$8m + 4n = 80 \quad (3)$$

and do $(3) - (2)$:

$$5m = 30 \Rightarrow \underline{\underline{m = 6}}$$
$$\Rightarrow \underline{\underline{n = 8}}.$$