Paper 5

ADVANCED LEVEL
BIOLOGY 9700

t-test

* t-test is primarily used to compare the means of two sets of data.

* It can help us determine if there is a significant difference in the means of the two sets of data

* We calculate the t-value and use the table of t-values to determine that.

To calculate the t-value .....

* To calculate the t-value, we need the;

1. Sample size (n_1 and n_2) of the two sets of data
2. Means ($\bar{x}_1$ and $\bar{x}_2$) of the two sets of data
3. standard deviations (s_1 and s_2) of the two sets of data.

$\bar{x} = 24$

data1 : 10 5 45 50 0

data2 : 41 43 45 47 50

$\bar{x} \approx 45$

measure of the spread of the data around the mean value

Formulae

t-value

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

degree of freedom

$$= n_1 + n_2 - 2$$

Interpretation of t-value

* If the calculated t-value is greater than the critical value at $p = 0.05$, (implies $p < 0.05$)

  → there is a significant difference in the means of the two sets of data

  → difference is NOT due to chance

  (null hypothesis rejected)

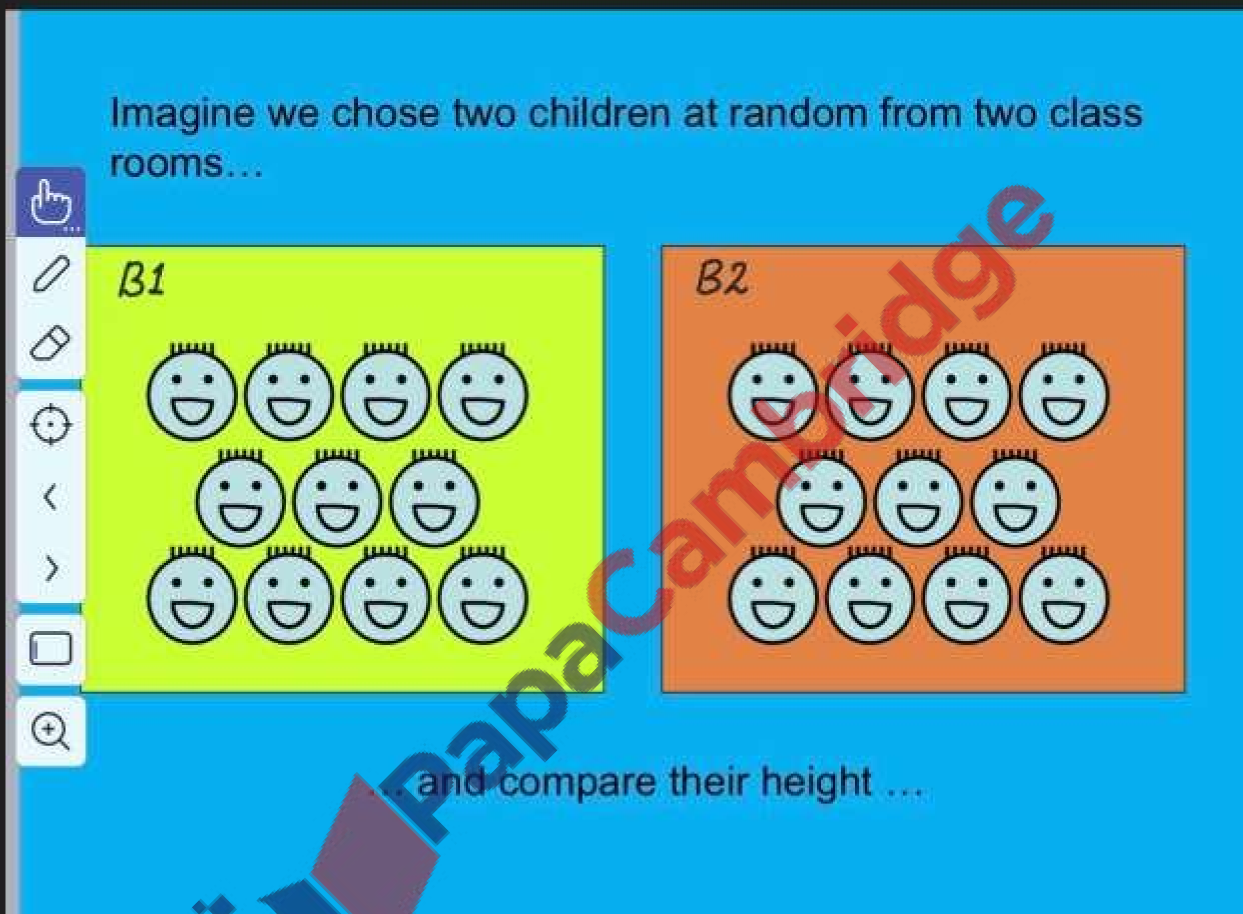
* If the calculated t-value is lesser than the critical value at $p = 0.05$, (implies $p \geq 0.05$)

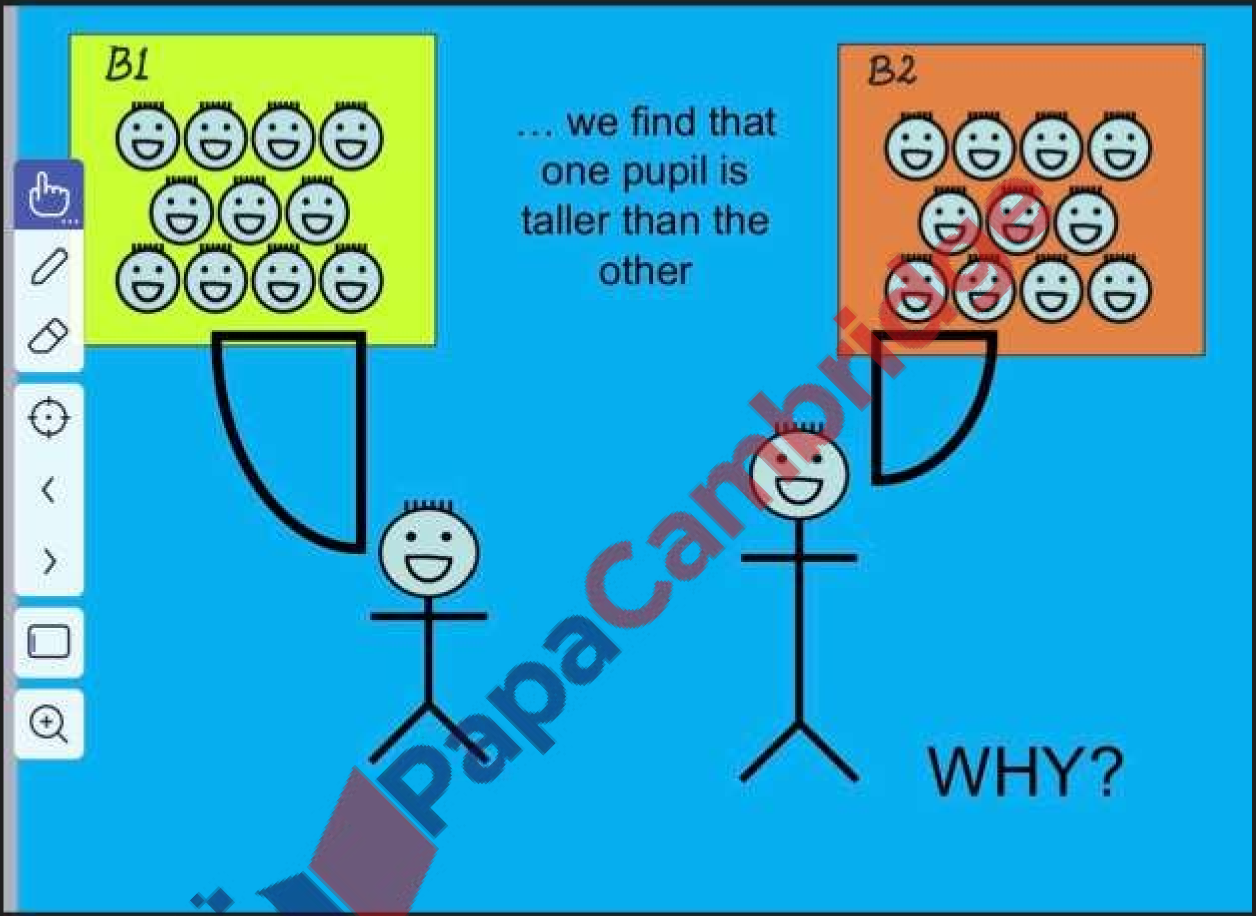
  → there is NO significant difference in the means of the two sets of data

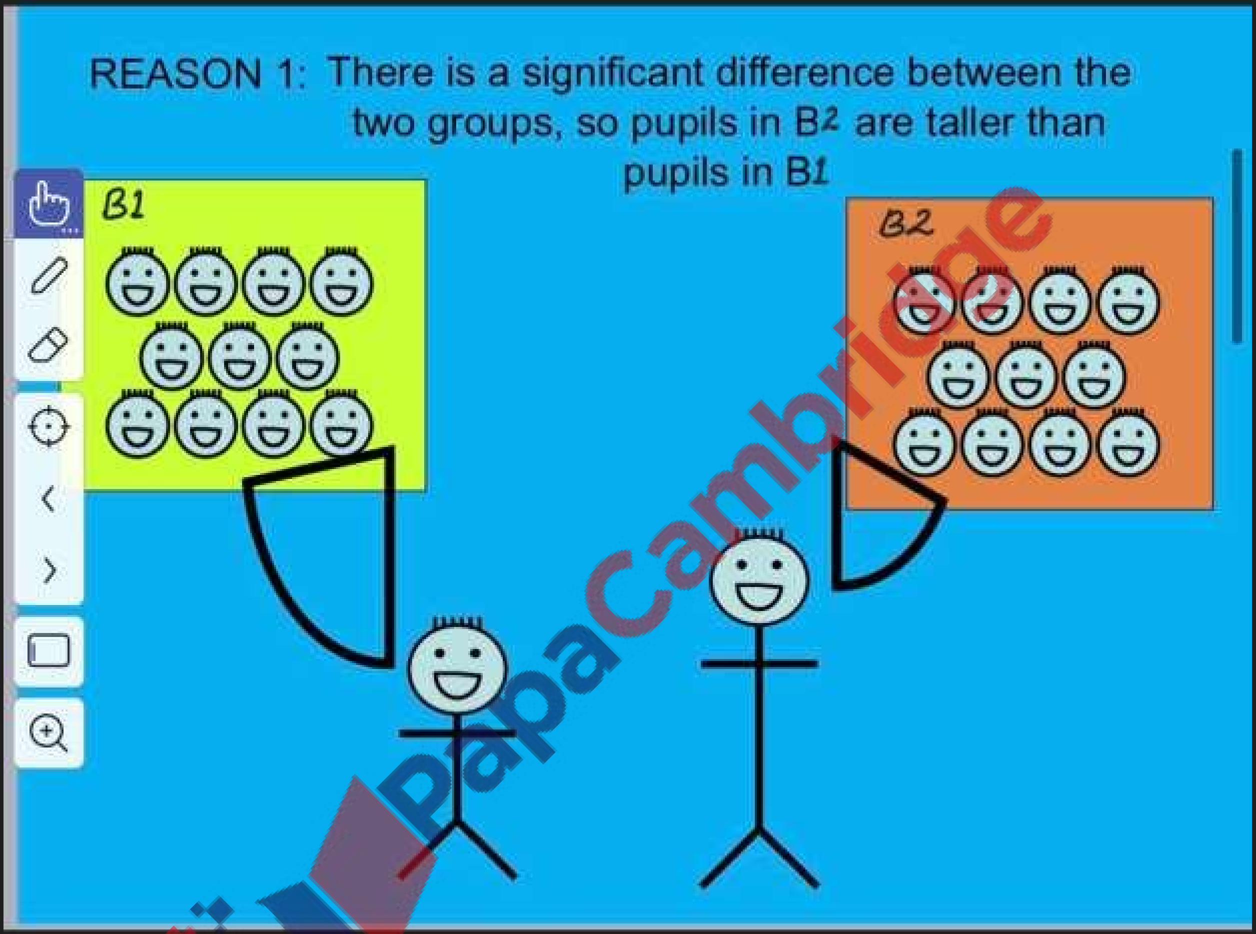
  → difference in results is just due to CHANCE

  (null hypothesis accepted)


Let's try an example ....


Imagine we chose two children at random from two class rooms…
B1
B2
… and compare their height …

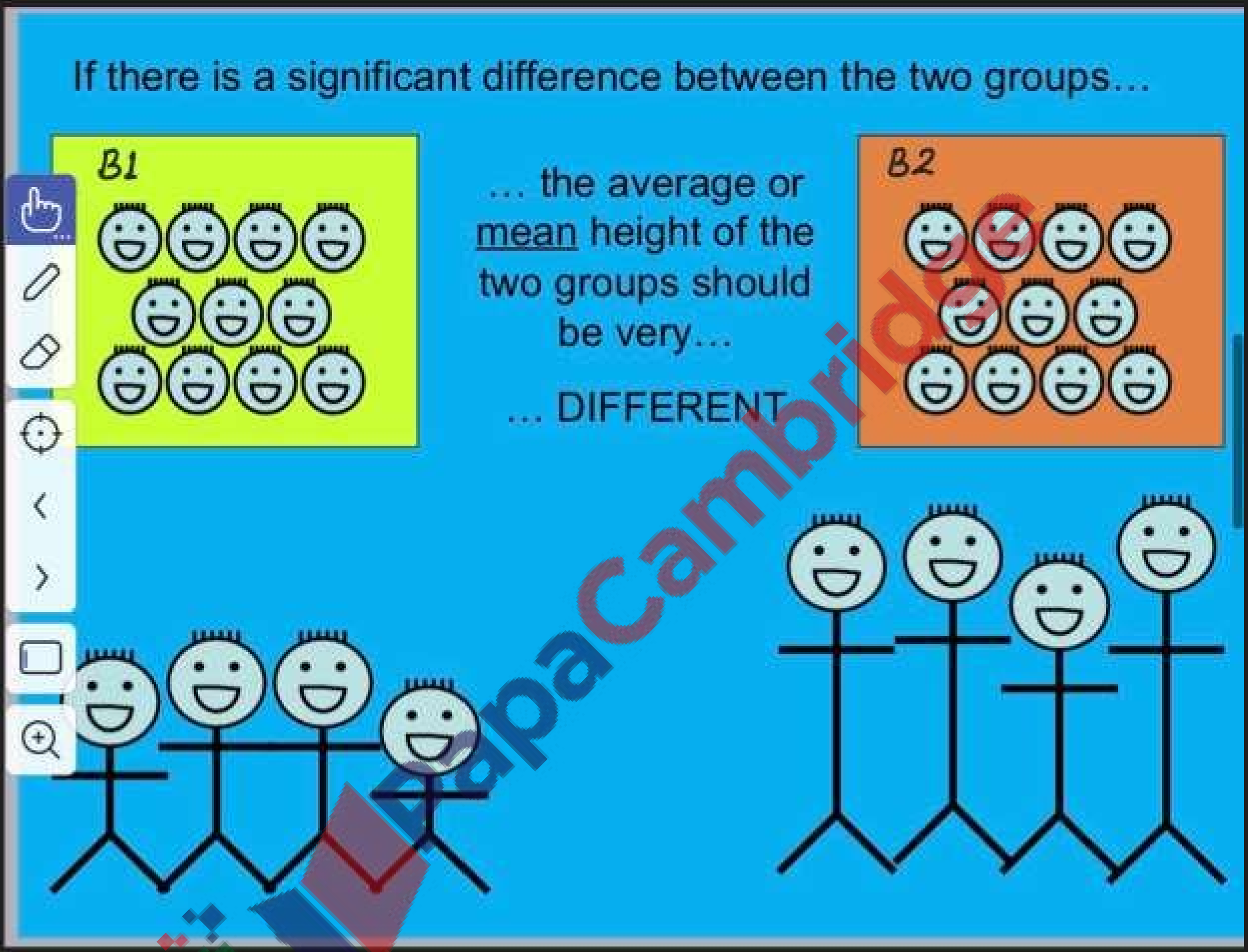

B1
... we find that one pupil is taller than the other
B2
WHY?

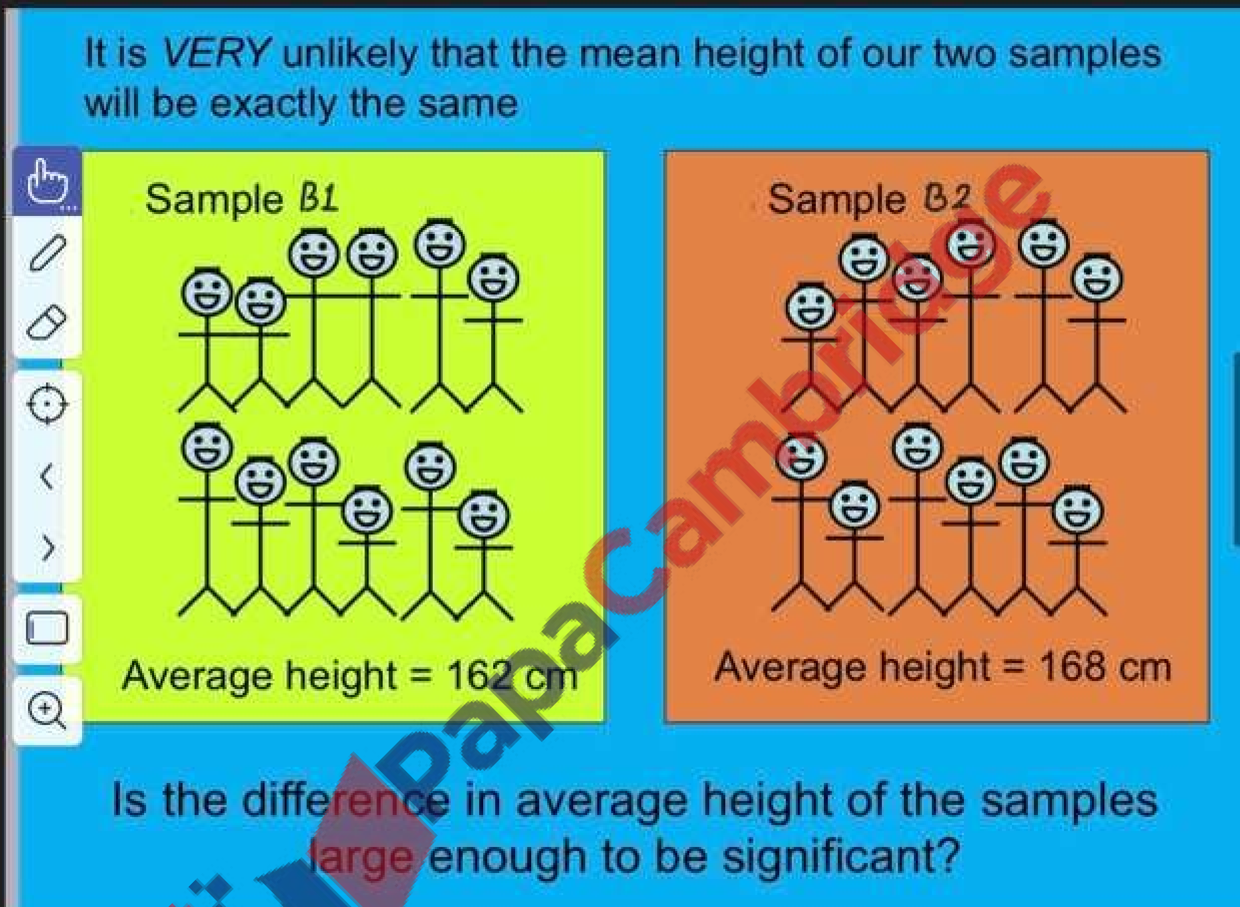

REASON 1: There is a significant difference between the two groups, so pupils in B2 are taller than pupils in B1
B1
B2


REASON 2: By chance, we picked a short pupil from B1 and a tall one from B2
B1
B2

How do we decide which reason is most likely?

Record the height of more STUDENTS!!!


If there is a significant difference between the two groups…
B1
… the average or mean height of the two groups should be very…
… DIFFERENT
B2


It is VERY unlikely that the mean height of our two samples will be exactly the same
Sample B1
Average height = 162 cm
Sample B2
Average height = 168 cm
Is the difference in average height of the samples large enough to be significant?

Student's *t*-test

The Student's *t*-test compares the averages and standard deviations of two samples to see if there is a significant difference between them.

We start by calculating a number, *t*

t can be calculated using the equation:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where:

- $\bar{x}$**1** is the mean of sample 1
- s**1** is the standard deviation of sample 1
- n**1** is the number of individuals in sample 1
- $\bar{x}$**2** is the mean of sample 2
- s**2** is the standard deviation of sample 2
- n**2** is the number of individuals in sample 2

Worked Example: Random samples were taken of pupils in B1 and B2

Their recorded heights are shown below…

	Students in B1					Students in B2				
Student Height (cm)	145	149	152	153	154	148	153	157	161	162
	154	158	160	166	166	162	163	167	172	172
	166	167	175	177	182	175	177	183	185	187

Step 1: Work out the mean height for each sample

B1: $\overline{x}1$ = 161.60 B2: $\overline{x}2$ = 168.27

Step 2: Work out the difference in means

$\overline{x}2 - \overline{x}1 = 168.27 - 161.60$

$= 6.67$

Step 3: Calculate

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{(7.86 + 9.19)} = 4.13$$

Step 4: Calculate t (Step 2 divided by Step 5)

$$t = \frac{6.67}{4.13} = 1.62$$

Step 5: Work out the number of degrees of freedom

d.f. = n1 + n2 – 2 =15 + 15 – 2 = 28

Step 6: Find the critical value of *t* for the relevant number of degrees of freedom

Use the 95% (p=0.05) confidence limit

Critical value = 2.048

Our calculated value of t is below the critical value for 28d.f., therefore, there is no significant difference between the mean height of students in samples from B1 and B2

Degrees of freedom	Significance level 20% (0.20)	10% (0.10)	5% (0.05)	2% (0.02)	1% (0.01)	0.1% (0.001)
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.043	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646

1 A student noticed that the leaves on a plant growing close to a wall had two sorts of leaves. The leaves next to the wall were in the shade and looked different from the leaves on the side away from the wall that were exposed to the sun. The length of the internodes on the stem also looked different.

The student decided to investigate the differences by measuring some features of 30 leaves and internodes from each side of the plant.

Fig. 1.1 shows the leaf shape

Fig. 1.2 shows an internode

Fig. 1.1

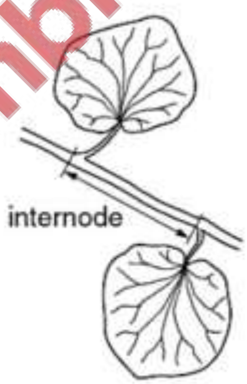


Fig. 1.2

Table 1.1 shows the student's results.

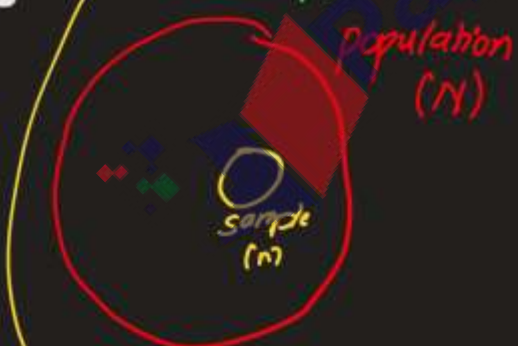
Table 1.1

	shaded leaves	exposed leaves
mean internode length / mm	23 ± 4	15 ± 3
mean surface area of leaves / mm^2	2750 ± 12	1800 ± 15
mean mass of leaves / mg	50 ± 8	60 ± 10
mean leaf surface area : leaf mass ratio	55 ± 9	30 ± 6
rate of water loss / mg mm^{-2} h^{-1}	50 ± 11	65 ± 12

(a) (i) State the independent variable being investigated.

light + intensity / exposure; [1]

$S_M = \frac{s}{\sqrt{n}}$



$(\bar{x} \pm \text{uncertainty})$

$\pm s$ $\pm S_M$ $\pm$ 95% CI $= \pm 2 S_M$

is a measure of how close the sample mean is to the actual mean value

(ii) Outline the procedures the student could use to obtain these results.

independent variable:

1.ref. to a systematic way of obtaining leaves
e.g. 3rd leaf from the apex / different heights / all from the same height / equal light exposure

dependent variables:

2. ref. to a method of measuring surface area (1)
e.g. draw round each leaf on grid or use transparent grid over leaf / measure diameter(s) of leaf
3. ref. to how surface area is calculated;
e.g. count squares / use formula $\pi r2$ (2)
4. ref. to a method of measuring mass; e.g. digital balance / scales
5. ref. to finding dry mass; e.g. sample leaves dried in oven until mass constant

(3) SA: mass

- ref. to a method of measuring internode length (4) either on the plant or a cut section from a plant; e.g.: by holding against a ruler / use string or cotton to mark distance measure with ruler
- (5) ref. to a method of measuring water loss; e.g.: use a potometer / weigh leaf / place leaf inside a plastic bag (to collect water)
- ref. to mean values of the whole sample;
- ref. to method of working out SA : mass ratio;
- ref. to calculating standard deviation

The student carried out *t*-tests for leaf surface area : leaf mass ratio and for internode length

The leaf surface area : leaf mass ratio gave the value $t = 12.6$

The formula for *t*-test is

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

(b) (i) Complete the calculation to find the value of *t* for the internode length.
Show your working.

(23±4)   (15±3)

$$t = \frac{23 - 15}{\sqrt{\frac{4^2}{30} + \frac{3^2}{30}}} = 8.76$$

(b) (i) Complete the calculation to find the value of t for the internode length.
Show your working.

$$t = \frac{23 - 15}{\sqrt{\frac{4^2}{30} + \frac{3^2}{30}}}$$

$$= \frac{(8)}{(0.9)} = \quad 8.9;$$

$t =$ 8.9 ....[3

Table 1.2 shows the critical values at $p < 0.05$ for the *t*-test.

Table 1.2

degrees of freedom	18	20	21	22	23	24	25	26	27	28	29	30	40	60	∞
critical value	2.10	2.09	2.08	2.07	2.06	2.06	2.06	2.06	2.05	2.05	2.04	2.04	2.02	2.00	1.96

The number of degrees of freedom is 58.

(ii) State how the number of degrees of freedom was calculated.

(30 – 1) + (30 – 1) = 58;

30 + 30 − 2 = 58 .... [1]

State and explain the meaning of these results.

ref. (both) calculated / t values (12.6 and 8.9) are greater than the critical value;

both results are significant / not due to chance / caused by another factor / light exposure;

[2]

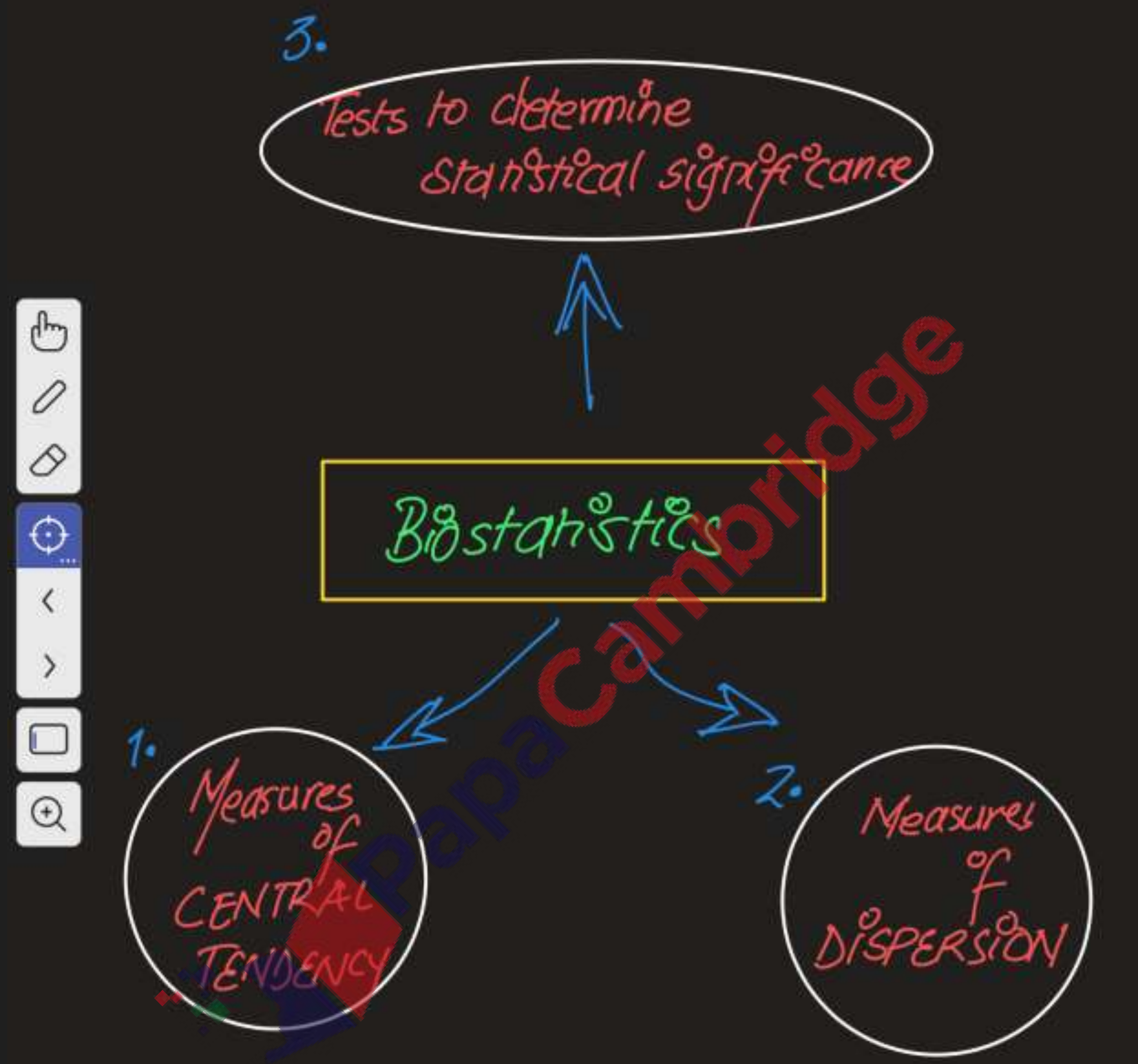
Paper 5

With

Mohammad Hussham Arshad, MD

ADVANCED LEVEL
BIOLOGY 9700


Biostatistics


3.
Tests to determine statistical significance
Biostatistics
1.
Measures of CENTRAL TENDENCY
2.
Measures of DISPERSION

① Measures of Central Tendency

a. Mean b. Mode c. Median

② Measures of Dispersion

a. Range b. Interquartile range

✓ c. Standard deviation (s)

d. Variance (s^2)

✓ e. Standard error (S_M)

✓ f. 95% confidence interval (CI)

③ Tests to determine statistical significance

a. χ^2 - test

b. t - test

c. S_M

d. 95% C.I.

Measures of Central Tendency

MEAN → average of the data

MEDIAN → central value(s) in the data

MODE → most common value(s) in the data

Basic Biostatistics

Measures of Central Tendency

- ***Mean***- not suitable as an 'average' value in the presence of extreme values ('outliers').
- ***Mode***- most frequent value(s) in a distribution. A distribution can have more than one mode.
- ***Median***- more suitable for skewed distributions

Practice Questions

1. Find the median of the set of numbers: 1,2,3,4,5,6,7,8,9 and 10.

a. 55

b. 10

c. 1

d. 5.5

2. Find the median of the set of numbers: 21, 3, 7, 17, 19, 31, 46, 20 and 43.

a. 19

b. 20

c. 3

d. 167

3, 7, 17, 19, 20, 21, 31, 43, 46

4. The following represents age distribution of students in an elementary class. Find the mode of the values: 7, 9, 10, 13, 11, 7, 9, 19, 12, 11, 9, 7, 9, 10, 11.

a. 7 (3)

b. 9 (4) ✓

c. 10 (2)

d. 11 (3)

5. Find the mode from these test results:

90, 80, 77, 86, 90, 91, 77, 66, 69, 65, 43, 65, 75, 43, 90.

a. 43

b. 77

c. 65

d. 90

6. Find the mode from these test results:

17, 19, 18, 17, 18, 19, 11, 17, 16, 19, 15, 15, 15, 17, 13, 11.

a. 15

b. 11

c. 17

d. 19

. Find the mean of these set of numbers:

00, 1050, 320, 600 and 150.

. 333

b. 444

. 440

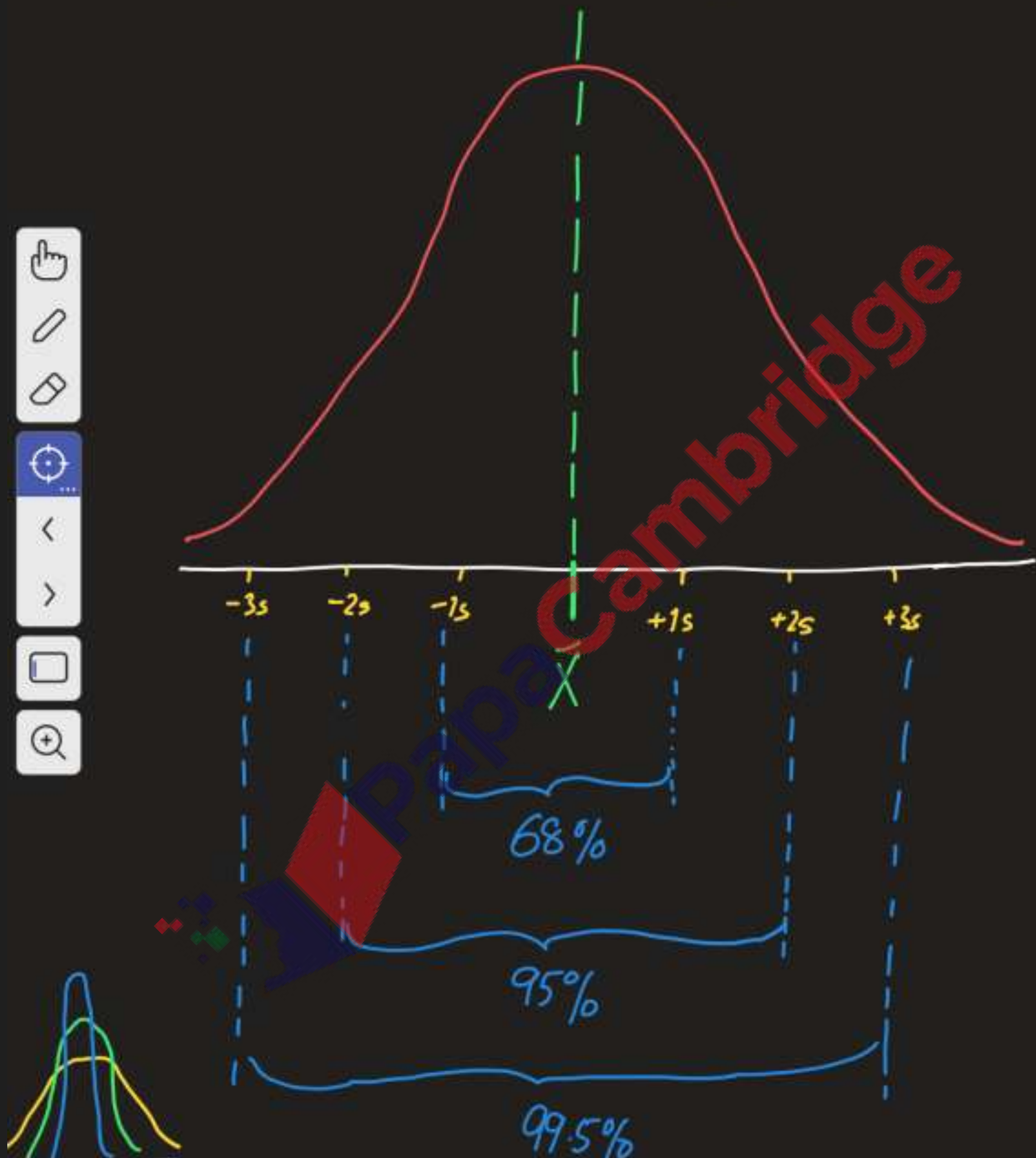
. 320

$$\bar{x} = \frac{100 + 1050 + 320 + 600 + 150}{5}$$

Normal Distribution

* a data is said to be normally distributed if its mean = mode = median

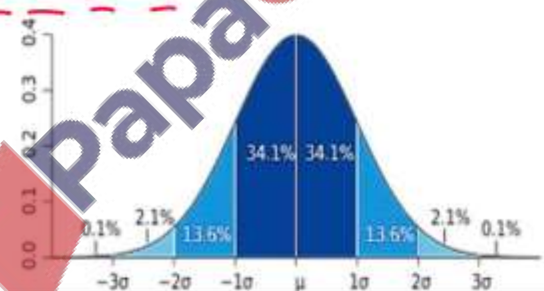
* Normally distributed data produce a symmetrical bell shaped curve



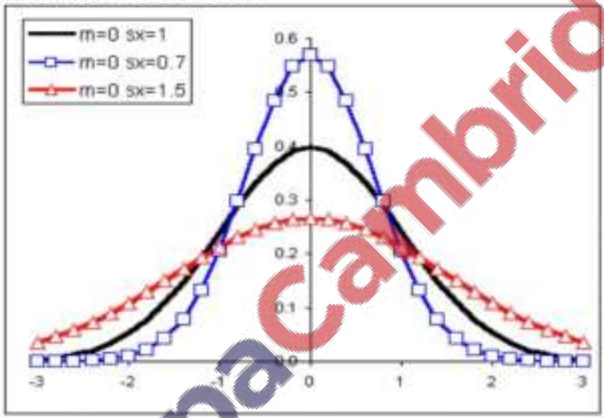
Normal Distribution

a) A normal distribution is a very important statistical data distribution pattern occurring in many natural phenomena, such as height, blood pressure, lengths of objects produced by machines, etc. Certain data, when graphed as a histogram (data on the horizontal axis, amount of data on the vertical axis), creates a bell-shaped curve known as a normal curve, or normal distribution.

b) Normal distributions are symmetrical with a single central peak at the mean (average) of the data. The shape of the curve is described as bell-shaped with the graph falling off evenly on either side of the mean. Fifty percent of the distribution lies to the left of the mean and fifty percent lies to the right of the mean.



c) The spread of a normal distribution is controlled by the standard deviation. The smaller the standard deviation the more concentrated the data.



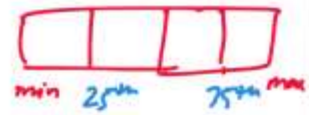
d) The mean and the median are the same in a normal distribution.


Measures of Dispersion
S
Sn
CI

Measures of Dispersion

- ***Range-*** difference between the maximum and minimum value in a distribution

- ***Interquartile range-*** difference between the upper (75th centile) and lower (25th centile) quartiles



- ***Standard deviation (σ):***

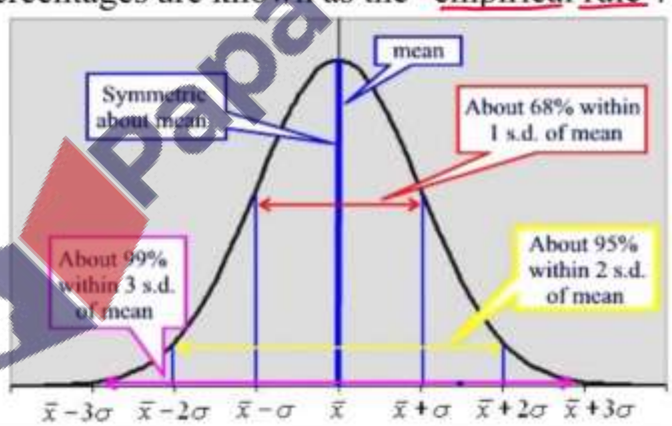
1. Represented by 's' and denoted by Greek letter σ (sigma)
2. Measure of how spread the data is across the mean
3. Larger the standard deviation, greater the spread across the mean
4. For a normally distributed data:

a) 68% of the distribution lies within one standard deviation of the mean.

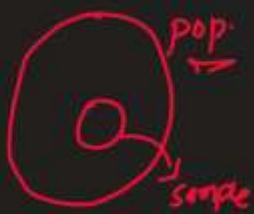
b) 95% of the distribution lies within two standard deviations of the mean.

c) 99.7% of the distribution lies within three standard deviations of the mean.

These percentages are known as the "empirical rule".



Calculating standard deviation (s)



$$s = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n-1}}$$

$\bar{x} \rightarrow$ mean

$n \rightarrow$ sample size

Q. Determine the standard deviation of the following data:

12, 6, 7, 3, 15, 10, 18, 5

Ans:

STEP 1 $\rightarrow$ determine $\bar{x} = 9.5$

STEP 2 $\rightarrow$ determine $\Sigma (x - \bar{x})^2$

x	$x - \bar{x}$	$(x - \bar{x})^2$
12	+2.5	6.25
6	−3.5	12.25
7	−2.5	6.25
3	−6.5	42.25
15	+5.5	30.25
10	+0.5	0.25
18	+8.5	72.25
5	−4.5	20.25

$\Sigma = 190$

STEP 3 → determine n → n = 8

STEP 4 → Calculate s

$$s = \sqrt{\frac{\Sigma (x-\bar{x})^2}{n-1}}$$

$$= \sqrt{\frac{190}{8-1}} = 5.21$$

Q. What's the range of this data?

Ans Range = maximum - minimum

= 18 - 3

= 15

- ***Variance (σ^2)-*** is the square of standard deviation

- ***Standard Error of the mean (S_M):***

a) Denoted by σ_M

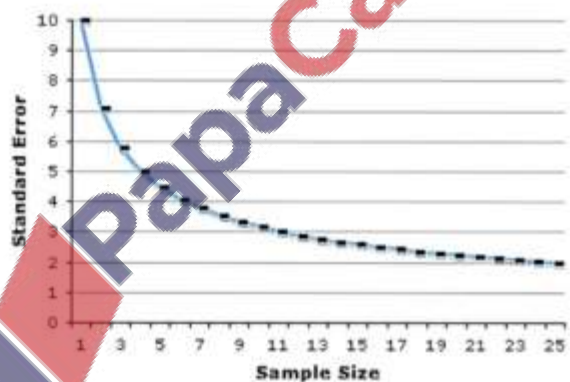
b) Represents how close the *data (sample)* mean is to the *actual(population)* mean value

c) …is a measure of how accurate the calculated mean value is

d) Lower the value of standard error (S_M), the more accurate is the mean value

e) Depends on the sample size and standard deviation of a given data

f) Standard error is proportional to $1/\sqrt{n}$, which implies that increasing the sample size (n) decreases the standard error.



$S_M = \frac{s}{\sqrt{n}}$

Comparing s and s_M ......

Statistical parameter

s

* spread of the data around the mean
* measure of reliability of the mean

• Smaller s
↓
more reliable mean

s_M

* how close the sample mean is to the actual mean
* measure of reliability/accuracy of the mean

• Smaller s_M
↓
more reliable/accurate the mean

95% confidence intervals (CI)

95% confidence interval

(95% C.I.)

$\pm 2 S_M$ → (mean $\pm 2 S_M$)

8 ± 2

$6 - 10$

$\pm 2s$ → (mean $\pm 2s$)

Q. What is meant by 95% confidence interval?

Ans. 95% confident that the mean lies within these limits. It's a measure of the reliability of the mean. The smaller the value of 95% C.I., the more reliable the mean.

Uncertainty in biological data (mean)

"Uncertainty" in biological data (mean)

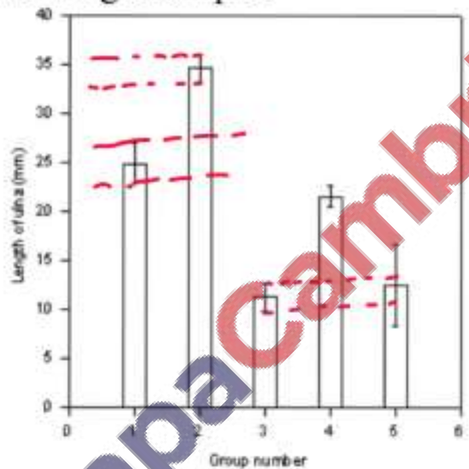
* can be represented in three ways

1. using standard deviation (s) → (mean ± s)
2. using standard error (S_M) → (mean ± S_M)
3. using 95% C.I → (mean ± 95% C.I.)

Graphical representation of uncertainty

- ***Error Bars***
  1. May be plotted using standard deviations or standard error of the mean
  2. Overlapping error bars: no significant difference between the data being compared
  3. Nonoverlapping error bars: significant difference between the data being compared

Let's study the following example:



Case: The histogram above shows 5 groups of individuals (1 till 5) in which the mean length of ulna was compared. The resulting error bars are plotted as shown above.

Interpretation: Using the Figure above, state:

1. Which group(s) has a significant difference in the mean length of ulna?
2. Which groups(s) have no statistically significant difference in their mean length of ulna?

Question

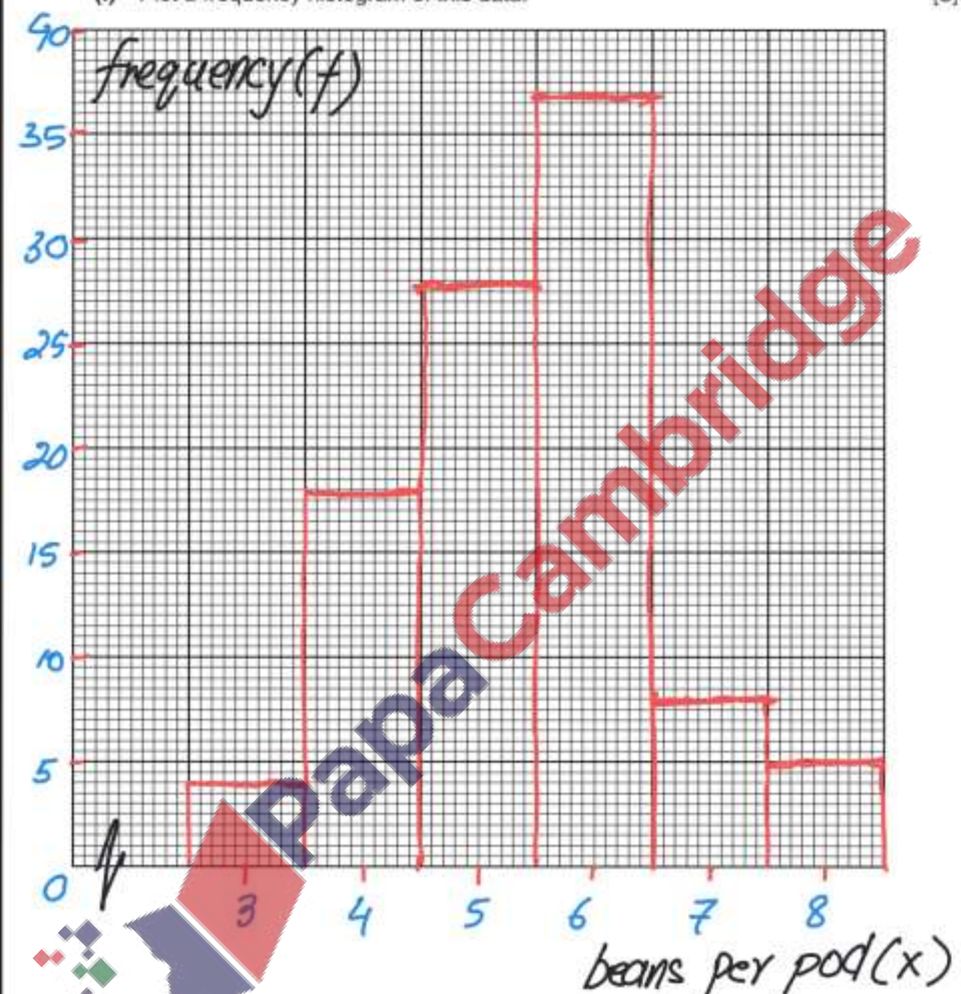
Q.

(a) 100 pods (n=100) from an inbred variety of bean were collected and the number of seeds in each pod counted. Table 1.1 shows the results of this investigation.

Table 1.1

number of beans per pod (x)	3	4	5	6	7	8
frequency (f)	4	18	28	37	8	5

(i) Plot a frequency histogram of this data. [3]



(ii) Complete Table 1.2 by calculating n, three values for fx and Σfx and putting the answers in the appropriate spaces on the table.

Table 1.2

number of beans per pod (x)	3	4	5	6	7	8	Total
frequency (f)	4	18	28	37	8	5	n = 100
fx	12	72	140	222	56	40	Σfx = 542

[1]

(iii) Use the formula to calculate the mean value ($\bar{x}$) of the number of seeds per pod.

$$\bar{x} = \frac{\Sigma fx}{n}$$

$\bar{x} = 542/100$

$\bar{x} = 5.42$ .... [1]

(iv) A student calculated the standard deviation (s) for this data.

The standard deviation, $s = 1.15$.

State what the standard deviation tells you about this investigation.

* the spread of the data around the mean

* spread for this data is (5.42 ± 1.15) [1]

(v) Use the formula to calculate the standard error (S_M) for this data.

$$S_M = \frac{s}{\sqrt{n}}$$

$S_M = \frac{1.15}{\sqrt{100}} = 1.15/10 = 0.115$

0.115 .... [1]

(b) Suggest an explanation to account for the different number of seeds in the pods of plants of the same genotype.

* might be growing in different environmental conditions

.... [1]

[Total : 8]

Q.

1 Fig. 1.1 shows a simple apparatus used by a student to measure the rate of respiration of yeast *Saccharomyces cerevisiae*.

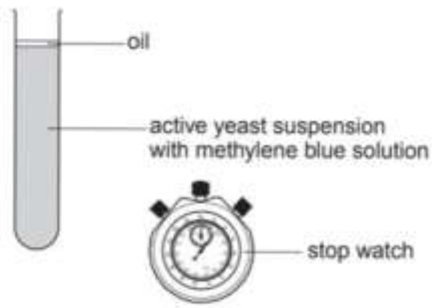


Fig. 1.1

Active yeast suspension is a pale cream colour. It is prepared by mixing dry yeast, glucose and water and leaving for 1 hour at 30 °C. The methylene blue solution acts as an electron acceptor and becomes colourless when reduced.

(a) **(i)** State how the dependent variable is measured in this experiment.

* time taken for blue colour to become colourless [1]

(ii) Explain why the layer of oil is needed.

* to prevent oxidation of methylene blue [1]

(iii) Outline how the student could use this apparatus to find the optimum temperature for the respiration of yeast.

....[7]

In a further investigation, the student tested the ability of yeast to use different sugars. Active yeast suspensions were mixed with 2% solutions of six different sugars. The yeast was allowed to metabolise the sugars at its optimum temperature and the carbon dioxide released was collected for a 10 minute period.

Table 1.1 shows the student's results.

Table 1.1

volume of carbon dioxide in 10 mins/cm^3											
monosaccharides						disaccharides					
glucose (glu)		fructose (fru)		galactose (gal)		sucrose (glu + fru)		maltose (glu + glu)		lactose (glu + gal)	
1	2.0	1	5.0	1	0.1	1	3.0	1	1.4	1	0.3
2	2.2	2	3.8	2	0.3	2	2.6	2	1.7	2	0.4
3	2.4	3	4.6	3	0.2	3	3.6	3	1.3	3	0.6
mean	2.2	mean	4.5	mean	0.2	mean	3.1	mean	1.5	mean	0.4

(b) Suggest an explanation for these results.

* yeast respires fructose, glucose, sucrose & maltose
* it has enzymes to metabolise these sugars
* yeast preferentially respires fructose
* sucrose has a high rate b/c it's a disaccharide made of fructose

....[3]

Based on these observations, the student made a hypothesis:

The yeast will form more cells when provided with fructose than with glucose.

Table 1.2 shows the results of counting cell samples from the active yeast suspension that had been supplied with fructose or glucose and left at the optimum temperature for 30 minutes. Three samples were taken from each suspension and the cells counted using a microscope slide with a grid. Four counts were made from each sample and the number of cells per mm^3 calculated.

Table 1.2

	fructose number of cells per mm^3					glucose number of cells per mm^3				
	count 1	count 2	count 3	count 4	mean	count 1	count 2	count 3	count 4	mean
sample 1	52	75	62	56		53	55	48	54	
sample 2	58	66	71	46	62	45	52	51	52	51
sample 3	65	61	68	64		53	53	42	54	

The student then calculated the standard error for these results.

The standard deviation for fructose = 8.11 = 8 cells

The standard error for fructose = 2.30 = 2 cells

The formula for standard error is: $S_M = \frac{s}{\sqrt{n}}$ s = standard deviation, n = number of samples

(c) (i) Complete the calculation to find the value of S_M for glucose.

Show your working. State your answer to the nearest whole cell.

$S_M = \frac{4}{\sqrt{12}}$ = 1.16

fructose vs glucose

$\bar{x}$ 62 51

s 8 4

S_M 2 1

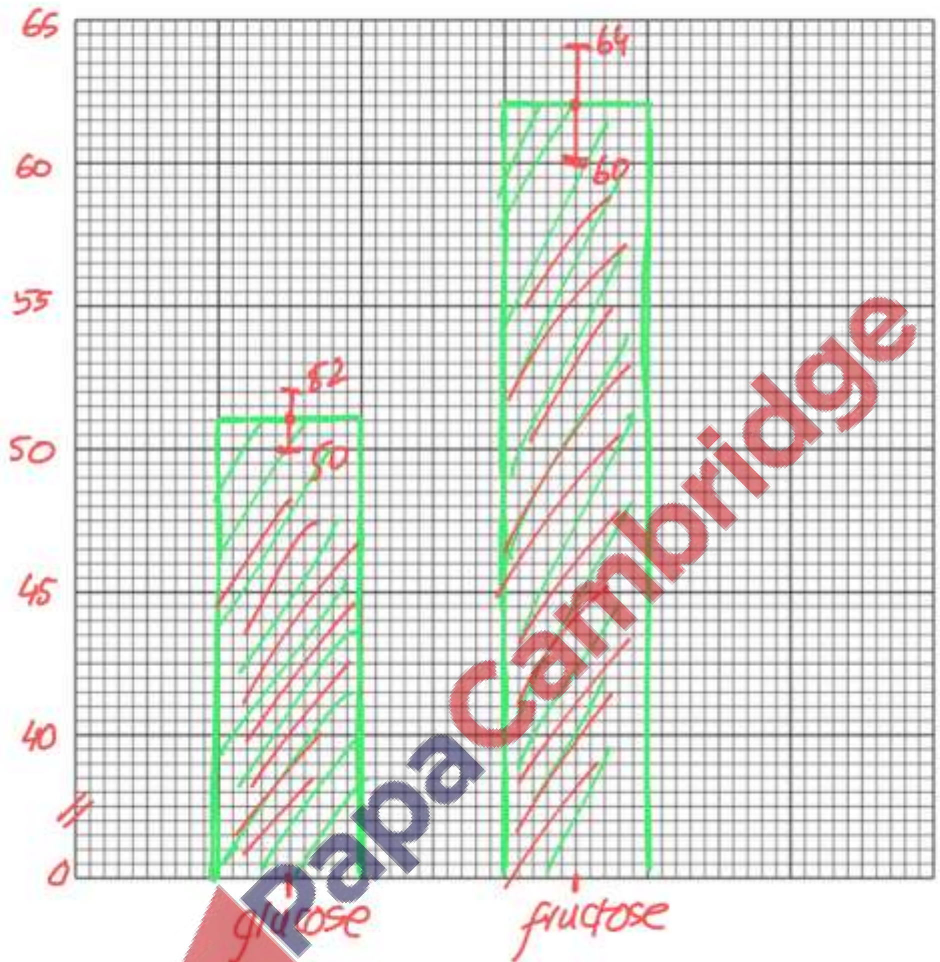
S_M = 1 cell [3]

(ii) State what standard deviation shows.

* spread of the data around the mean

* larger the value the less reliable the mean [2]

(iii) Use the grid to plot the data in Table 1.2 to show the difference in the population size of yeast supplied with fructose and yeast supplied with gluc
Include error bars.



[3]

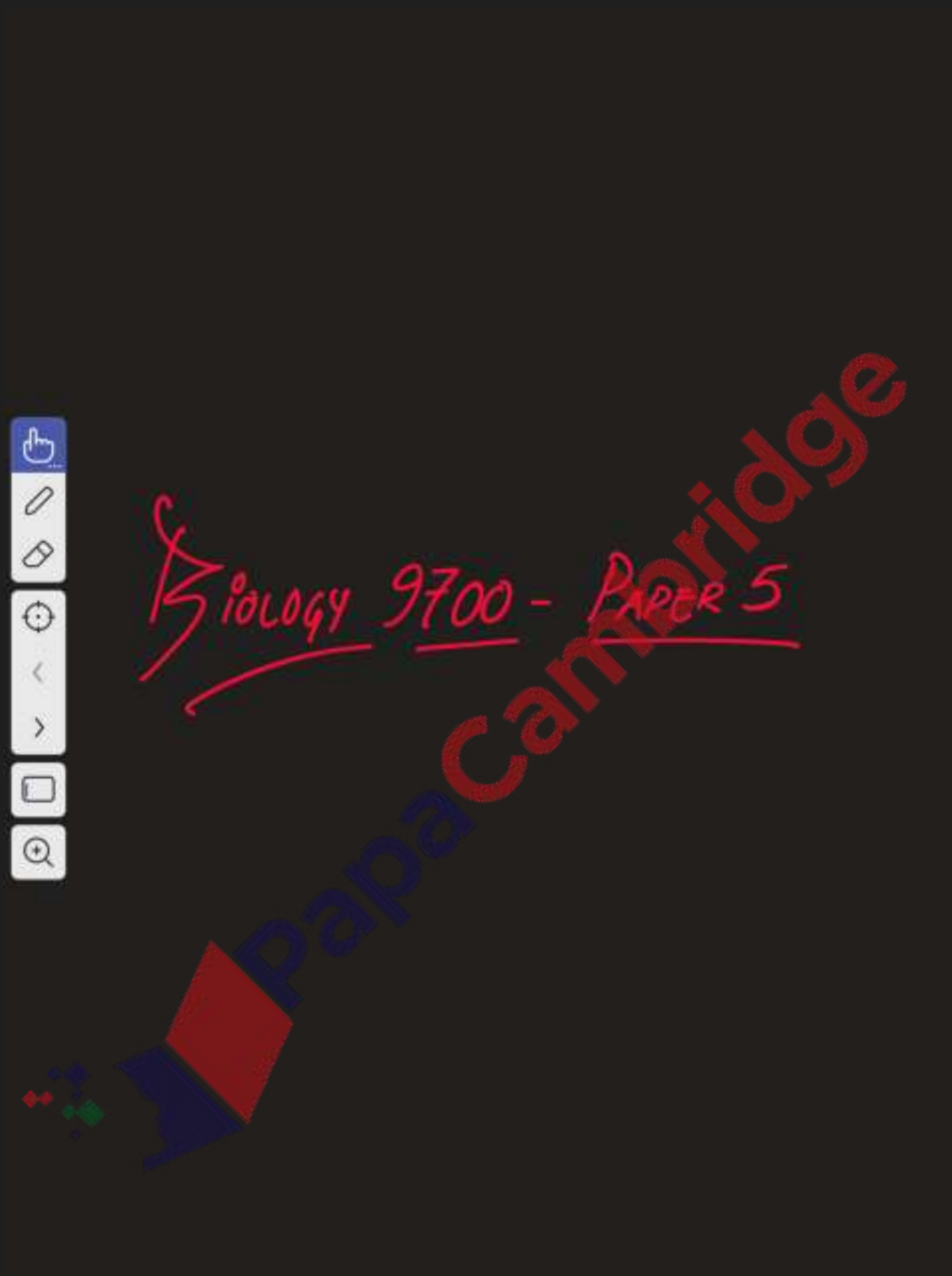
(iv) State whether the data support the hypothesis. Give a reason for your answer.

yes → standard error bars do NOT overlap

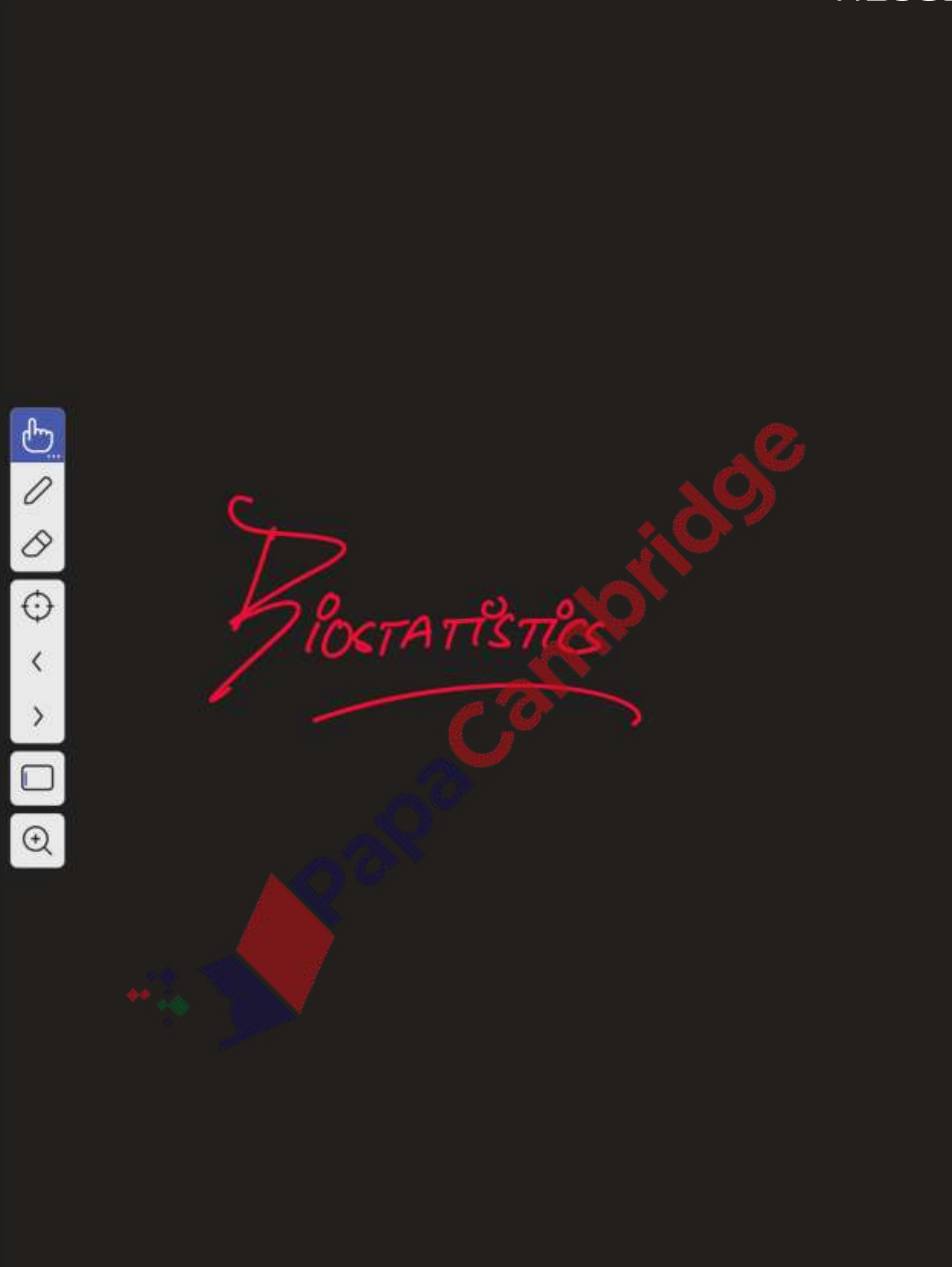
.... [1]

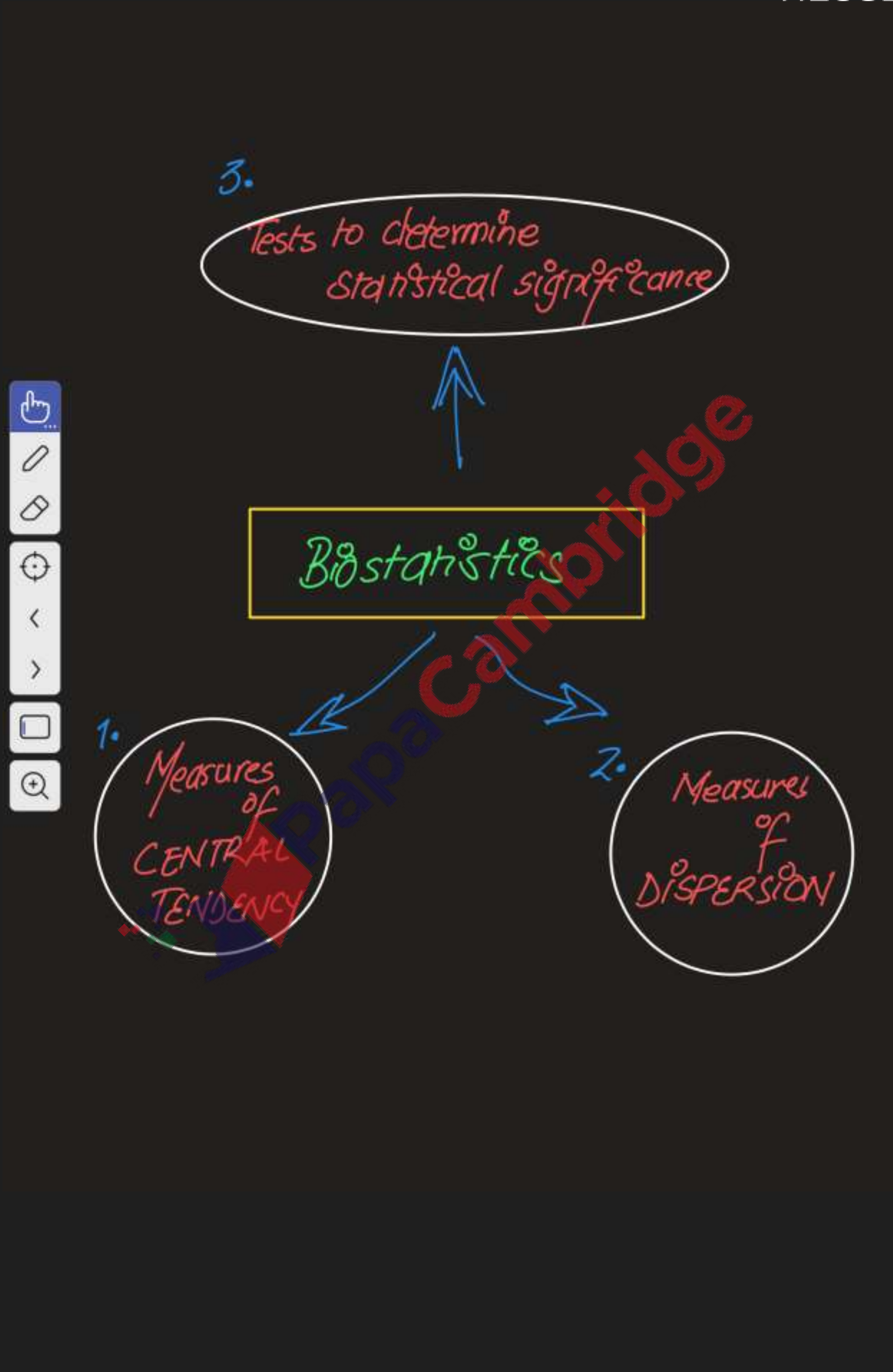
[Total: 21]

Paper 5



ADVANCED LEVEL
BIOLOGY 9700


Biostatistics


3.
Tests to determine statistical significance
Biostatistics
1.
Measures of CENTRAL TENDENCY
2.
Measures of DISPERSION

(1) Measures of Central Tendency

a. Mean b. Mode c. Median

(2) Measures of Dispersion

a. Range b. Interquartile range

c. Standard deviation (s)

d. Variance (s^2)

e. Standard error (s_M)

f. 95% confidence interval

(3) Tests to determine statistical significance

a. χ^2 - test

b. t - test

c. S_M

d. 95% C.I.

Uncertainty in biological data (mean)

"Uncertainty" in biological data (mean)

* can be represented in three ways

1. using standard deviation (s) → (mean ± s)
2. using standard error (S_M) → (mean ± S_M)
3. using 95% C.I → (mean ± 95% C.I.)


Question

Q.

The students also investigated the effect grazing had on the height of one particular species of plant. Their hypothesis was:

The mean height of the plant is greater in the ungrazed grassland than the grazed grassland.

(c) State the independent and the dependent variables in this investigation.

independent variable ....grazed and ungrazed grassland....

dependent variable ....mean height of the plant.... [1]

(d) Table 1.1 shows the results of their investigation.

Table 1.1

sample number	height of plant/mm	
	grazed area	ungrazed area
1	586	858
2	549	873
3	526	864
4	589	901
5	545	847
6	538	862
7	573	864
8	549	879
9	604	864
10	611	888
mean	567	870
mode	549	864
median	561	864

(i) Complete Table 1.1 by writing the values of the mode and median for the ungrazed area. [1]

(ii) Use the information and formula below to calculate the standard error for these results.

Give your answers to 3 significant figures.

$S_M = \frac{s}{\sqrt{n}}$

S_M = standard error

s = standard deviation

n = sample size (number of observations)

grazed area: $s = 29.5$ $S_M = s/\sqrt{n} = 29.5/\sqrt{10} = 9.35$

ungrazed area: $s = 15.7$ $S_M = s/\sqrt{n} = 15.7/\sqrt{10} = 4.97$

standard error, grazed area = ± 9.35

standard error, ungrazed area = ± 4.97 [2]

Standard error is used to calculate 95% Confidence Intervals (CI).

The values for the grazed area are 548.3 mm to 585.7 mm.

(iii) Use the formula below to calculate the confidence intervals for the **ungrazed** area.

$$95\%\ CI = \text{mean} \pm 2\ S_M$$

Show your working.

$95\%\ CI = 870 \pm 2(4.97)$

$= 870 \pm 9.94$

ungrazed area 860.1 mm to 879.9 mm [2]

(iv) State what information is gained by calculating the confidence intervals.

* it implies that you're 95% confident that the mean lies within this range
* it's a measure of reliability of the mean

[2]

Q.

(c) Some local communities hold annual fishing trips where they are encouraged to catch as many lionfish as possible.

In 2010 and 2011, 16 boats took part in these fishing trips. The mean number of lionfish per boat and the 95% confidence intervals were calculated.

Fig. 2.2 shows the results.

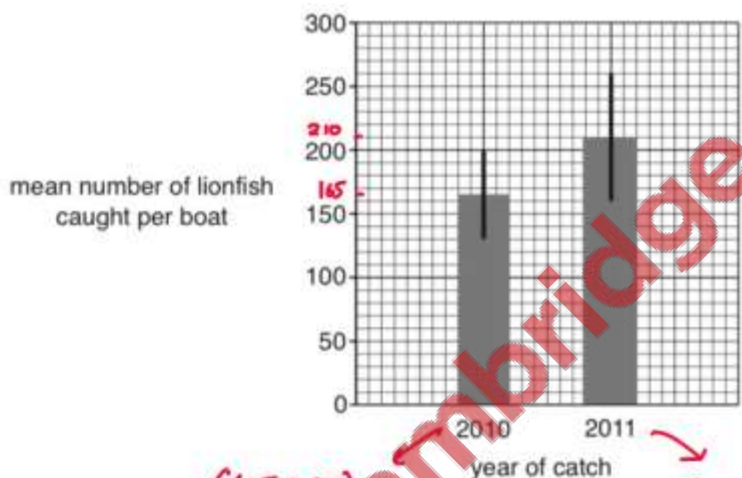


Fig. 2.2

The 95% confidence intervals are calculated from $2 \times S_M$.

Standard deviation can be calculated by rearranging the equation for standard error shown in Fig. 2.3.

$$S_M = \frac{s}{\sqrt{n}}$$

where S_M = standard error

s = standard deviation

n = number of boats

Fig. 2.3

2010

$S_M = s/\sqrt{n}$

$2 \times S_M = 35$

$S_M = 35/2 = 17.5$

$17.5 = s/\sqrt{16}$

$s = 17.5 \times 4 = 70$

Use the data in Fig. 2.2 and the equation in Fig. 2.3 to calculate the standard deviation of the mean number of lionfish caught per boat in 2010.

$2 \times S_M = 35$

$S_M = 17.5$

$17.5 = s/\sqrt{16}$

$s = 70$

standard deviation for 2010 = ...70.... [3]

Q.

2 Parkinson's disease occurs when nerve cells in part of the brain die, resulting in less secretion of a neurotransmitter called dopamine.

Dopamine is involved in the control of muscle movement and in emotional responses. As a result, common symptoms of Parkinson's disease are muscle twitching and stiff muscles. Other symptoms can include depression, memory loss and problems with sleeping.

The disease is progressive, so over time the symptoms become worse.

(a) Scientists collected data on the number of males and females with Parkinson's disease in specific age groups in the USA.

Fig. 2.1 shows these data presented as number of people with Parkinson's disease per 100 000.

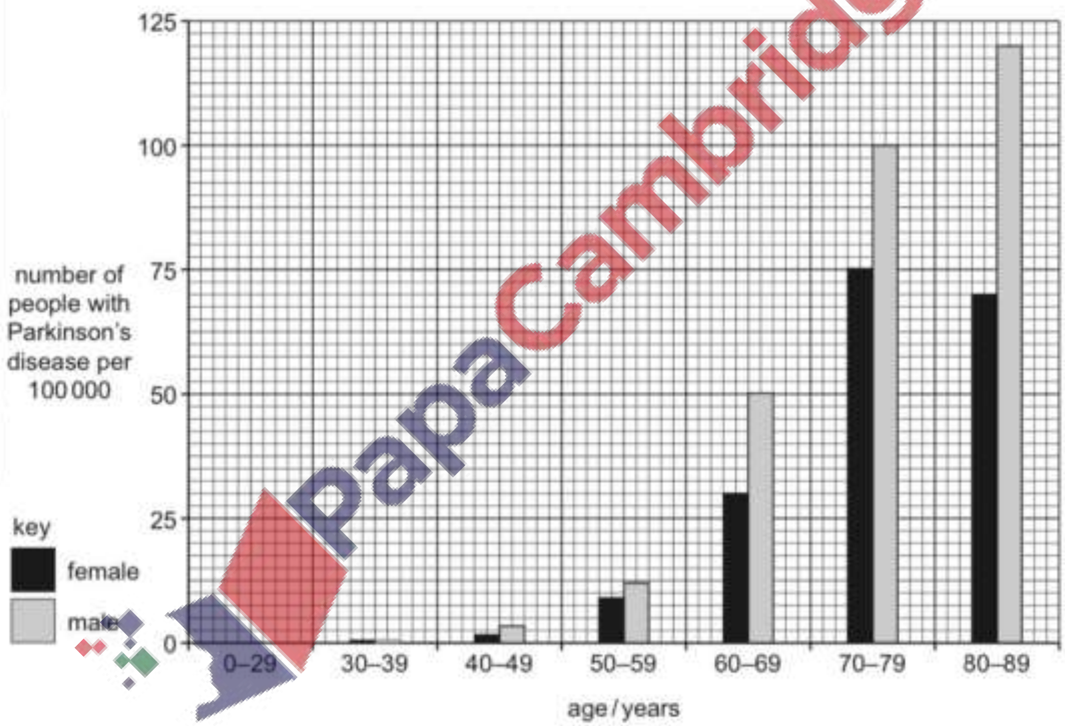


Fig. 2.1

(i) State what conclusions can be made from Fig. 2.1 about Parkinson's disease in the USA.

* the number of people with Parkinson's disease (per 100000) increase with increasing age
* the no. of males is greater than the no. of females in each age category
* the steepest rise in the no of cases is between the age categories 60-69 and 70-79

[3]

(ii) A student suggested that the Spearman's rank correlation could be used to test the relationship between age and the occurrence of Parkinson's disease.

Give a reason why it is possible to use the Spearman's rank correlation to analyse the data shown in Fig. 2.1.

* b/c the data can be ranked

[1]

(iii) State a null hypothesis for this test.

* there is no correlation between the age and the occurence of Parkinson's Disease

[1]

(b) There is currently no cure for Parkinson's disease although a range of treatments can be used to relieve the symptoms. Drugs are given that are absorbed by brain cells and converted to dopamine. As the disease progresses, more brain cells die and so less of the drug is absorbed. As a result, the effectiveness of these drugs decreases.

Research into the use of neural stem cells (NSC) as a more effective treatment for Parkinson's disease is being carried out. One study was carried out to test the hypothesis that:

> The use of a Chinese herbal drug increases the differentiation of NSC into neurones that produce dopamine.

In this study, 48 healthy male rats of similar mass were randomly divided into four groups. Each rat was kept in a separate cage and supplied with food and water.

Fig. 2.2 shows the experimental design of the study.

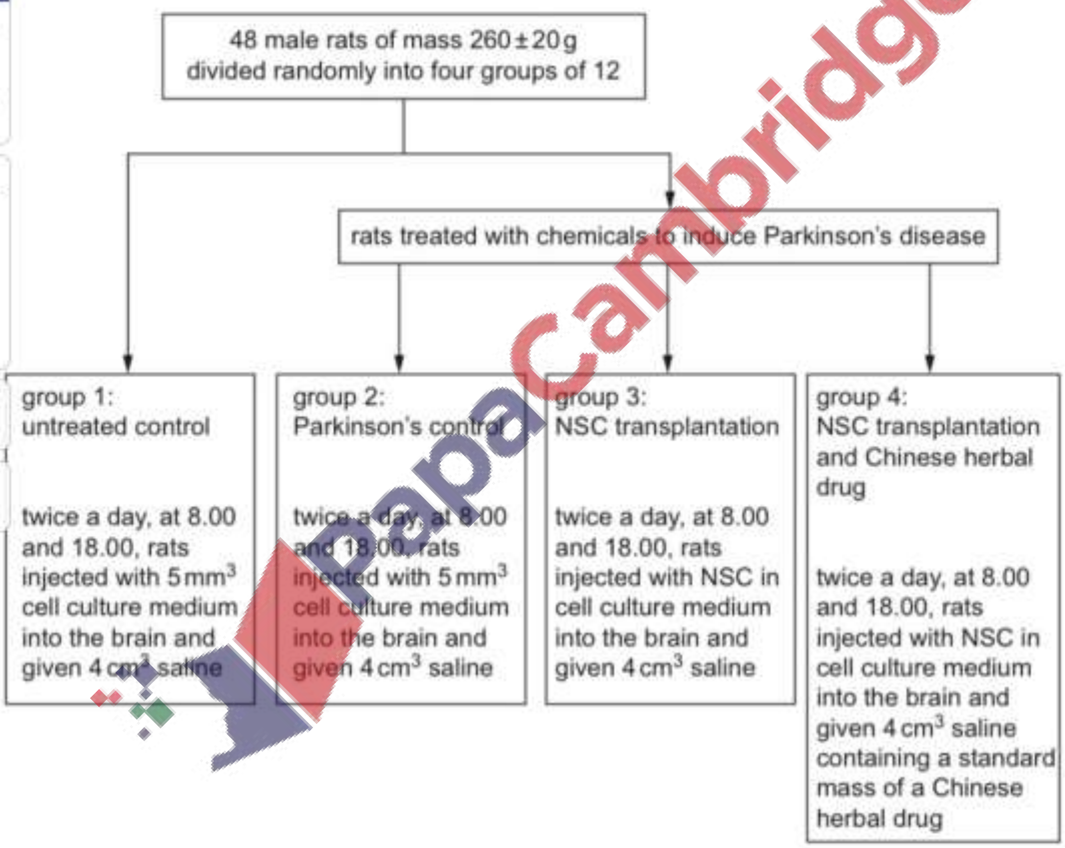


Fig. 2.2

(i) Identify **two** variables that have been standardised in this study.

* number of rats in each group
* mass of rats (± 20g)

[2]

(ii) Suggest **one** variable that should have been standardised in this study but for which there is no information about how this was done.

* age of the rats [1]

(c) At 7, 14 and 28 days after the start of the study, a sample of four rats from each group was tested to find the dopamine concentration in the brain.

The results are shown in Table 2.1.

Table 2.1

time after treatment / days	mean dopamine concentration ± standard deviation (*s*) / nmol per g of brain tissue			
	group 1: untreated control	group 2: Parkinson's control	group 3: NSC transplantation	group 4: NSC transplantation and Chinese herbal drug
7	59.8 ± 3.3	33.5 ± 5.1	50.2 ± 2.8	86.8 ± 4.7
14	60.0 ± 4.4	31.6 ± 7.6	49.9 ± 4.8	81.8 ± 27.1
28	60.1 ± 3.0	31.9 ± 3.7	39.5 ± 2.6	46.5 ± 1.1

State **two** advantages of calculating standard deviations (*s*) in studies of this type.

* spread of the data around the mean
* can be used to plot the error bars

[2]

(d) The researchers concluded that, in rats, the use of a Chinese herbal drug increases the differentiation of NSC into neurones that produce dopamine.

With reference to all the information about this study, give **one** piece of evidence that supports this conclusion and **one** piece of evidence that suggests this conclusion is not valid.

evidence that supports this conclusion the mean dopamine conc. is significantly higher in group 4 than group 3

evidence that suggests this conclusion is not valid only 12 rats were used in each group

[2]

[Total: 12]

Q.

2 Biodiversity is important in maintaining the stability of an ecosystem. Biodiversity can be reduced by the introduction of new species to an ecosystem.

The red quinine tree, *Cinchona pubescens*, was introduced to the Galapagos Islands in the 1940s. By 2010, it had covered 110 000 hectares of Santa Cruz, one of the larger Galapagos islands.

(a) A group of scientists studied the impact of the red quinine tree on the plant biodiversity of Santa Cruz.

Line transects were used to study an area of 32 hectares in the hills of Santa Cruz.

(i) Suggest **three** variables that the scientists needed to standardise in this study.

1 length of the transect

2 time of the year

3 size of the quadrats used

[3]

In addition to the line transects, the scientists set up 14 sample plots in the study area.

(ii) Suggest a method that could be used for randomly selecting the position of the plots.

* random number generator to generate the co-ordinates

[1]

(b) The scientists decided to use Simpson's Index of Diversity to calculate the plant biodiversity of the study plots.

State what data they needed to collect to calculate Simpson's Index of Diversity.

* the number (n) of organisms of each individual population

[1]

(c) Over a 7-year period the scientists measured the:

- ground covered by red quinine trees
- ground covered by other plant species
- ground **not** covered by plants.

Fig. 2.1 shows the results.

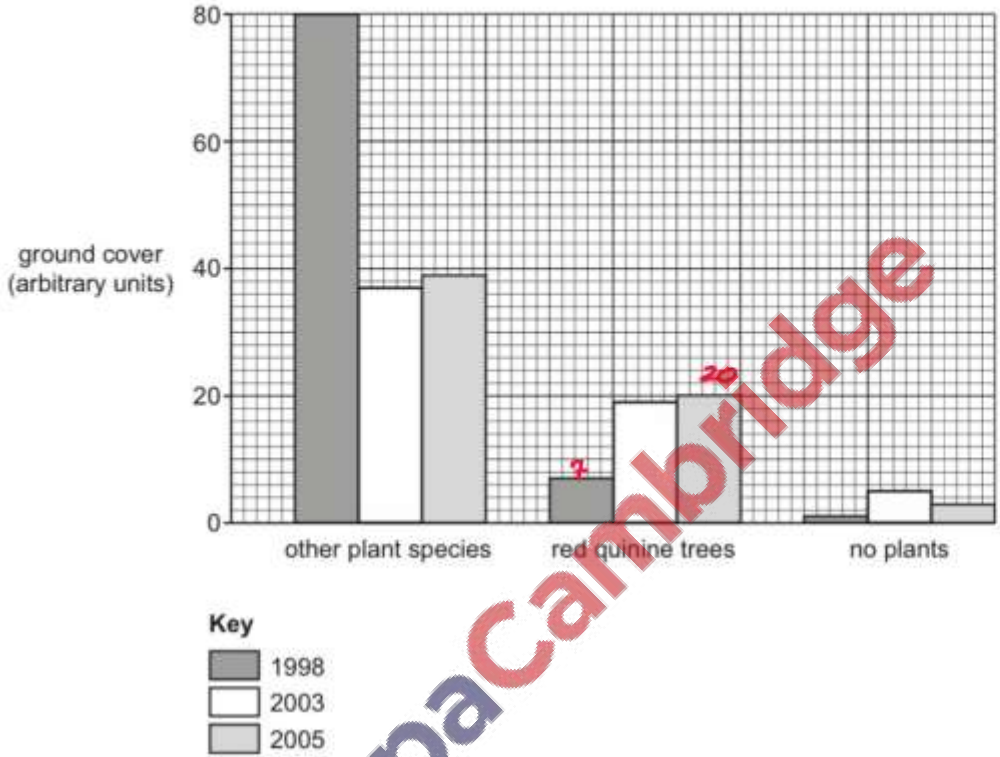


Fig. 2.1

Calculate the percentage change in the ground covered by red quinine trees in the plots from 1998 to 2005.

$$\% \text{ change} = \frac{20-7}{7} \times 100 = \frac{13}{7} \times 100$$

$$= 185.7$$

185.7 % [1]

(d) The scientists suggested the hypothesis:

The presence of red quinine trees in the hills of Santa Cruz causes a decrease in biodiversity.

The scientists carried out statistical tests on their studies of species diversity.

The probability values (p) from the results of the statistical tests are shown in Table 2.1.

Table 2.1

	value of p
decrease in species diversity 1998–2003	< 0.001
decrease in species diversity 2003–2005	> 0.05

Evaluate the data in Fig. 2.1 and Table 2.1 and discuss the extent to which the data supports or does **not** support this hypothesis.

SUPPORT

* increase in red quinine cover b/w 1998-2003 decreases the cover of other plant species
* increase in red quinine cover b/w 1998-2003 also increases the ground cover with no plants
* thus significant ($p < 0.001$) decline in biodiversity b/w 1998-2003

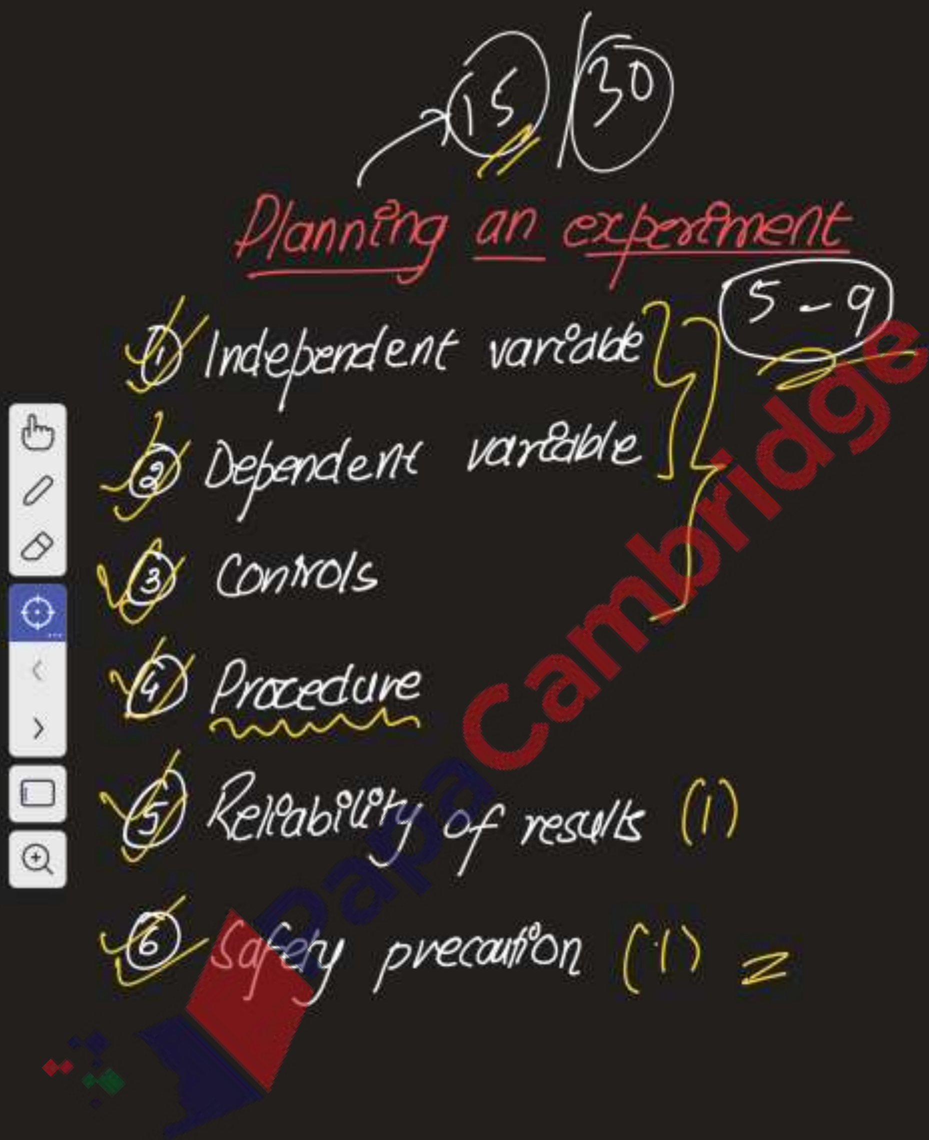
DOES NOT SUPPORT

* there is no data regarding animal biodiversity

[4]

[Total:10]

Paper 5


15
30
Planning an experiment
5 - 9
1 Independent variable
2 Dependent variable
3 Controls
4 Procedure
5 Reliability of results (1)
6 Safety precaution (1)
2

ADVANCED LEVEL
BIOLOGY 9700

1 (a) The opening and closing of stomata involves the movement of potassium ions into and out of guard cells. Opening and closing of stomata is influenced by a number of environmental factors, for example light and temperature.

A student investigated the effect of potassium chloride (KC*l*) on the opening of stomata.

The student was provided with:

- 500 cm^3 of 250 mmol dm^{-3} KC*l* solution
- freshly picked leaves from a plant that had been kept in the dark and a high concentration of carbon dioxide for an hour. This ensured that all the stomata were closed.

Strips of leaf tissue were obtained by cutting a leaf into sections as shown in Fig. 1.1.

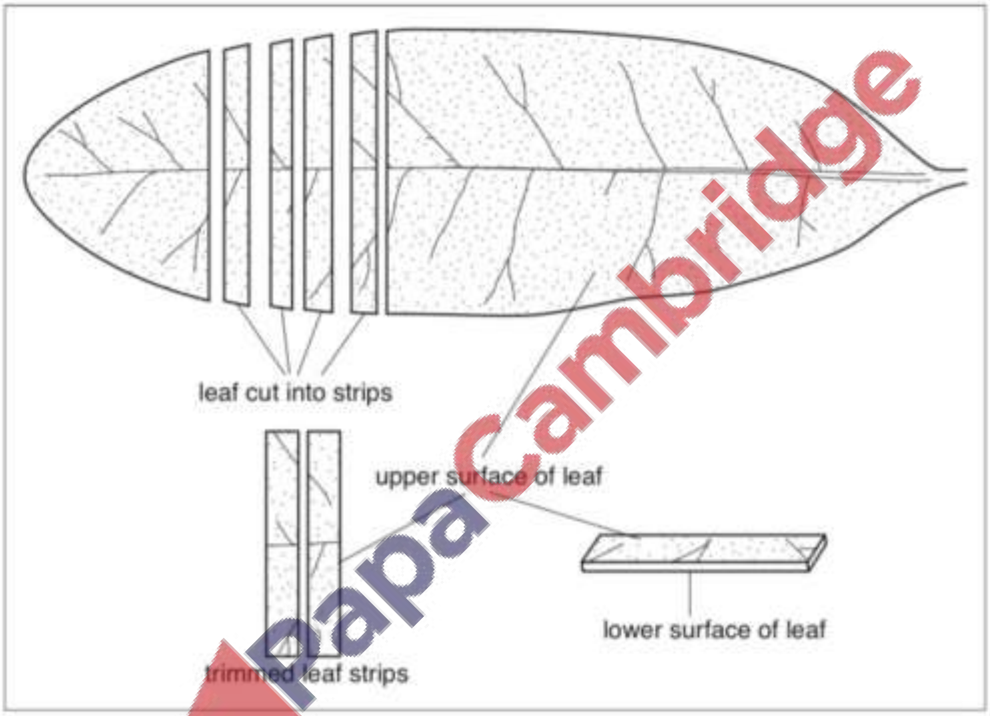


Fig. 1.1

The student floated three strips of leaf tissue in each of a range of buffered potassium chloride solutions for 2 hours and then recorded the number of open stomata.

(i) Identify the independent and dependent variables in this investigation.

independent variable concentration of KCl

dependent variable number of stomata open/closed

[2]

(ii) The student used the 250 mmol dm^{-3} KCl solution to make 100 cm^3 of four other concentrations by reducing the concentration by 50 mmol dm^{-3} each time.

Describe a procedure that the student could use to prepare these four concentrations.

[KCl] / mmol dm^{-3}	Vol. of [KCl] = 250 / cm^3	Vol. of distilled water / cm^3
200	80	20
150	60	40
100	40	60
50	20	80

* the volumes are measured using a measuring cylinder to create four conc. using simple dilution

* ensure proper mixing with a stirrer.

[3]

(b) (i) Suggest a hypothesis that the student could test about the effect of KCl on the opening and closing of stomata.

the higher the concentration of KCl, the greater the number of stomata open

[1]

(ii) Describe a method that the student could use to investigate the effect of different concentrations of KCl on the opening of stomata.

The description of your method should be detailed enough for another person to follow and should **not** repeat the details from **(a)(ii)** of how to dilute the 250 mmol dm^{-3} solution of KCl.

* place the strips into all the KCl solutions in petri dishes
* petri dishes must be kept in the dark
* mount each strip on a microscope slide and count the number of stomata
* Ensure that the same magnification is used to count the number of stomata in each strip

controls

* Maintain a constant temp. throughout the experiment by controlling room temp.
* petri dishes should be covered to minimize evaporation
* Use a scalpel and a metre rule to ensure all the strips are cut to the same length and width
* Take a minimum of 3 counts from each strip and take the mean (1)
* Take care while cutting strips. It's a low risk experiment (1)

(c) The student also tested the hypothesis:

The more light the wider the stomata open.

- Eight leaves from young plants that had been kept in the dark for 24 hours were covered by metal foil.
- A fluorescent lamp of fixed intensity was placed 10 cm from the plant. The metal foil was removed from the leaves.
- Two leaves were removed at the start of the experiment and three epidermal strips were made from each leaf. An epidermal strip is made by peeling the epidermis from a leaf as a single layer.
- The diameter of the stomatal aperture of five of the stomata with the widest aperture on each strip was measured.
- At one hour intervals two more leaves were removed and the same procedure repeated.

Fig. 1.2 shows stomata at different stages of opening.

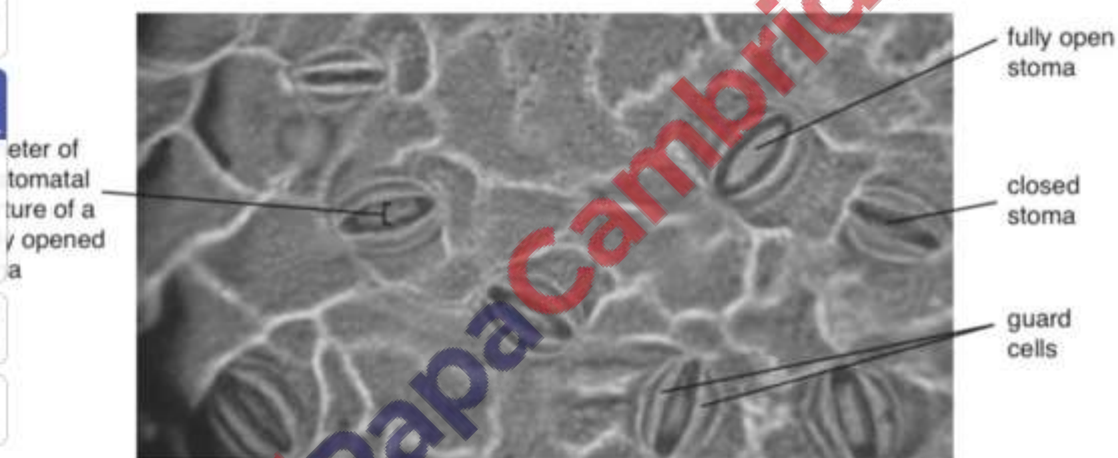


Fig. 1.2

(i) Outline how the student could find the actual diameter of a stomatal aperture.

* use an eyepiece graticule to measure the stomatal aperture
* calibrate the eyepiece graticule using a stage micrometer

.... [2]

Table 1.1 shows the results of the student's experiment.

Table 1.1

time / min	diameter of stomatal aperture / µm														
0 (control)	0.5	0.1	0.2	0.3	0.4	0.1	0.5	0.2	0.3	0.3	0.1	0.2	0.2	0.2	0.4
60	0.9	1.1	1.0	1.3	1.2	1.8	1.5	0.8	0.2	1.3	1.1	0.8	1.0	1.9	0.9
120	1.9	2.4	2.6	2.6	2.5	2.2	2.8	2.4	2.4	3.9	2.6	2.3	2.5	2.2	2.7
180	4.1	4.8	4.2	4.0	5.7	4.7	3.9	4.1	5.5	4.5	4.3	4.0	3.1	4.1	4.3

(ii) **On Table 1.1**, draw circles around **two** values that are anomalous. [1]

(iii) The student calculated the mean diameter of the stomatal apertures and the rate at which the diameter of the stomatal apertures increased.
Table 1.2 shows some of these calculations.

Table 1.2

time / min	mean diameter of stomatal apertures / µm	rate of increase of diameter of stomatal apertures / µm min^{-1}
0	0.3	
60	1.2	0.015
120	2.5	0.022
180	4.6	0.035

Complete Table 1.2 by calculating the rate of increase of the diameter of the stomatal apertures between 120 minutes and 180 minutes.

Space for working

[1]

(iv) The experimental procedure described in **(c)** could be criticised for poor technique in obtaining results.

Suggest how the procedure could be modified to improve the quality of these results.

* measure all the stomata per epidermal strip

* measure at shorter time intervals

* measure in more leaf strips

[3]

(d) The experimental procedure used in **(c)** is not completely valid for the stated hypothesis:

The more light the wider the stomata open.

Suggest how this hypothesis could be modified to match the procedure described in **(c)**.

the longer the time for light exposure → the wider the aperture

[1]

[Total: 19]

1 A student used the internet to find information about caffeine and the possible effects it has on the body.

The student decided to test the hypothesis that:

Caffeine decreases the reaction time and increases the heart rate.

Table 1.1 shows the results of the student's research into the caffeine content of drinks.

Table 1.1

drink	cup of coffee	cup of tea	100 cm^3 energy drink	350 cm^3 cola
caffeine content / mg	80	40	30	45

The student also found that it takes about 45 minutes before caffeine from a drink is fully absorbed into the blood and about 4 to 5 hours for the effect to wear off.

After further research, the student decided:

- test subjects should not know whether they are having a drink with caffeine or without caffeine
- drinks with caffeine and without caffeine should look identical
- to measure the subject's reaction time using a computer programme that flashes a coloured light at random intervals. The subject presses a switch as soon as they see the light
- to measure the heart rate using a pulse meter attached to the first finger.

(a) Suggest why it is important that the test subjects should not know whether they are having a drink with caffeine or without caffeine.

test results could be affected by subject expectations

[1]

(b) (i) State the independent variable and the **two** dependent variables in this investigation.

independent variable *presence/absence of caffeine*

dependent variables *heart rate AND human reaction time*

[2]

(ii) Use the information given to describe a method by which the student could test their hypothesis using **one** of the drinks.

Your method should be detailed enough for another person to follow.

* induct a large number (more than 20) of test subjects
* subjects should have drinks with AND without caffeine
* subjects should NOT be able to distinguish between the two drinks
* test subjects should NOT have taken any caffeine for at least 5 hrs before the test
* each subject should be tested in isolation from others
* subject should be "at rest" while taking measurements
* take measurements (of heart rate & reaction time):
  → before giving the drink
  → 45 minutes after giving the drink
* same volume of drink to all subjects
* test subjects should have similar age and mass
* take 3 readings of each subject and calculate the mean
* ensure subjects are not allergic to caffeine

(c) In another experiment 10 subjects were each given a different concentration of caffeine. The reaction time was measured 5 times for each subject and a mean calculated.

Table 1.2 shows the results.

Table 1.2

subject	1	2	3	4	5	6	7	8	9	10
concentration of caffeine / $mg\,dm^{-3}$	0	40	60	80	100	120	140	160	180	200
mean reaction time / ms	355	343	340	321	300	305	288	252	242	204

Fig. 1.1 shows the graph produced by a computer programme for the data.

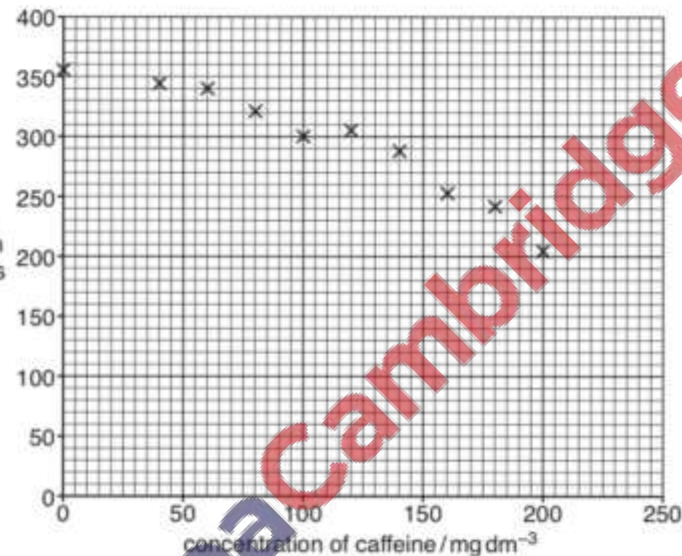


Fig. 1.1

Based on this graph the student decided to use a statistical test to find the strength of the correlation between the concentration of caffeine and the mean reaction time.

(i) State why Pearson's linear correlation test is suitable for the data.

* data is continuous
* data is normally distributed

[1]

(ii) The results of the statistical test gave Pearson's linear correlation, $r = -0.722$.

State what this value indicates about the relationship between the concentration of caffeine and the mean reaction time.

* negative linear correlation

[1]

Table 1.3 shows part of a table of critical values for Pearson's linear correlation test (r).

Table 1.3

number of pairs of data (n)	probability level (p)			
	0.10	**0.05**	**0.02**	**0.01**
1	0.988	0.997	0.9995	0.9999
2	0.900	0.950	0.980	0.990
3	0.805	0.878	0.934	0.959
4	0.729	0.811	0.882	0.917
5	0.669	0.754	0.833	0.874
6	0.622	0.707	0.789	0.834
7	0.582	0.666	0.750	0.798
8	0.549	0.632	0.716	0.765
9	0.521	0.602	0.685	0.735
10	0.497	0.576	0.658	0.708

(iii) Describe how the student calculated the degrees of freedom.

df = no. of pairs of data − 2
= 10 − 2 = 8

[1]

(iv) Describe how the student used the probability table to find out if the value for $r = 0.722$ is significant.

* use the values at $p=0.05$
* determine the critical value at $p=0.05$ for 8 degrees of freedom
* if the calculated r is greater than the critical value → r is significant

[3]

(d) Suggest **two** reasons why the method used for the investigation into the effect of caffeine concentration may have given results that are **not** reliable.

* only one person tested at each concentration

* the result of one person may be anomalous

.... [2]

(e) Effects of caffeine in the body include the promotion of acetylcholine release and the inhibition of acetylcholinesterase.

Suggest how this could account for the results in Table 1.2.

* concentration of acetylcholine remains high in synapses which makes the reaction time faster. [1]

[Total: 20]

1 The enzyme urease catalyses the breakdown of urea into carbonate ions and ammonium ions as shown in Fig. 1.1.

$$\underset{\text{urea}}{CO(NH_2)_2} + \underset{\text{water}}{2H_2O} \xrightleftharpoons{\text{urease}} \underset{\text{carbonate ion}}{CO_3^{2-}} + \underset{\text{ammonium ion}}{2NH_4^+}$$

Fig. 1.1

A solution of urea does not conduct electricity, but a solution of the two ions does conduct electricity.

An increase in ions increases conductivity, so the rate of reaction is proportional to the conductivity.

A meter measures conductivity in microsiemens per centimetre ($\mu S\,cm^{-1}$). The conductivity meter also records temperature.

K_m shows the affinity of an enzyme for its substrate. The lower the K_m the greater the affinity.
A student carried out an investigation to find the K_m of the enzyme urease at different temperatures.

The student:
- made $500\,cm^3$ of $0.2\,mol\,dm^{-3}$ solution of urea, molar mass $60.1\,g\,mol^{-1}$
- used serial dilution to make a total of five urea solutions from $0.2\,mol\,dm^{-3}$ stock solution
- used 1 g per $10\,cm^3$ urease solution
- used a conductivity meter to measure the initial rate of reaction at each temperature.

Fig. 1.2 shows the experimental set-up during the course of a reaction.

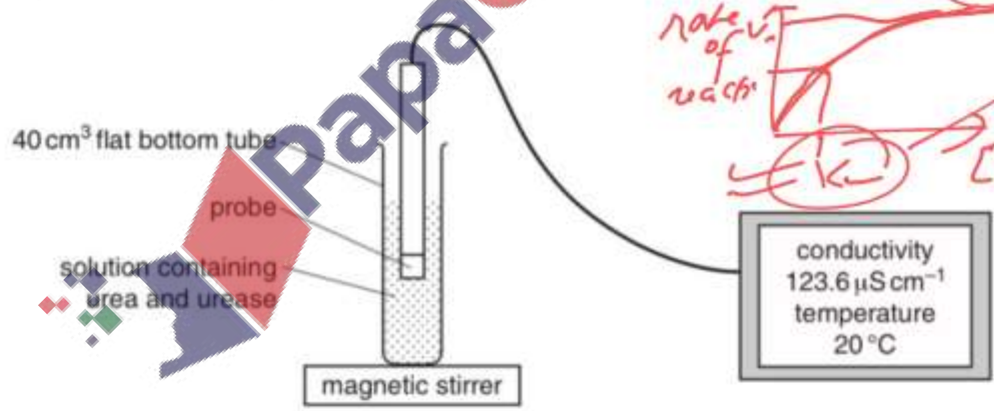


Fig. 1.2

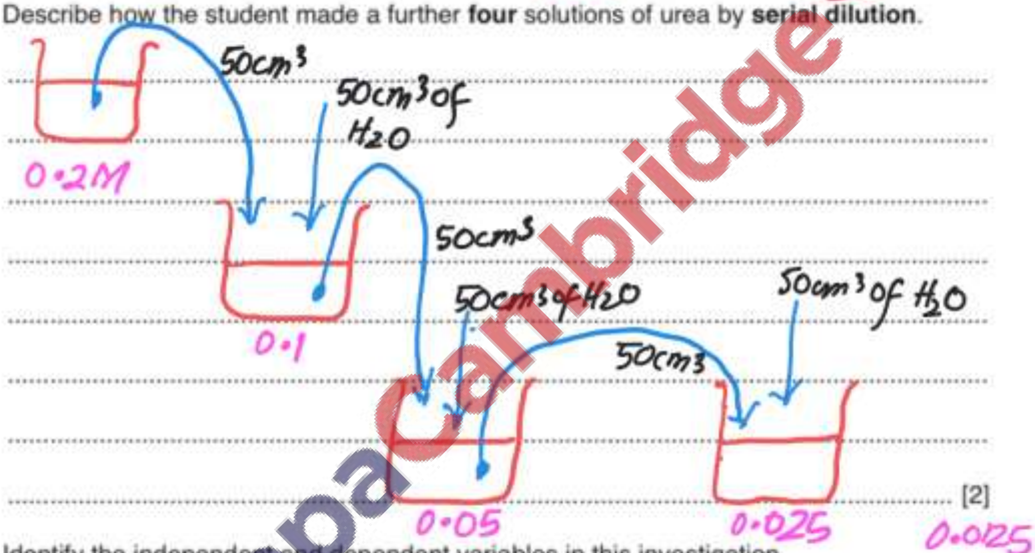
(a) (i) Describe how the student could make 500 cm^3 of 0.2 $mol\,dm^{-3}$ solution of urea.

* Weigh 6.01g of urea and
* add 500 cm^3 of distilled water to it

$c = n/V$

$0.2 = n/0.5$ $n = 0.1$ [2]

(ii) Describe how the student made a further **four** solutions of urea by **serial dilution**.



[2]

(b) (i) Identify the independent and dependent variables in this investigation.

independent temperature

dependent conductivity of enzyme and substrate [2]

(ii) Suggest a suitable control for this investigation.

* Substituting the active enzyme by a boiled enzyme for all experiments [1]

(iii) Describe a method the student could use to find the K_m value of urease at different temperatures. The solutions were made as described in **(a)(i)** and **(a)(ii)** and the apparatus shown in Fig. 1.2 was used.

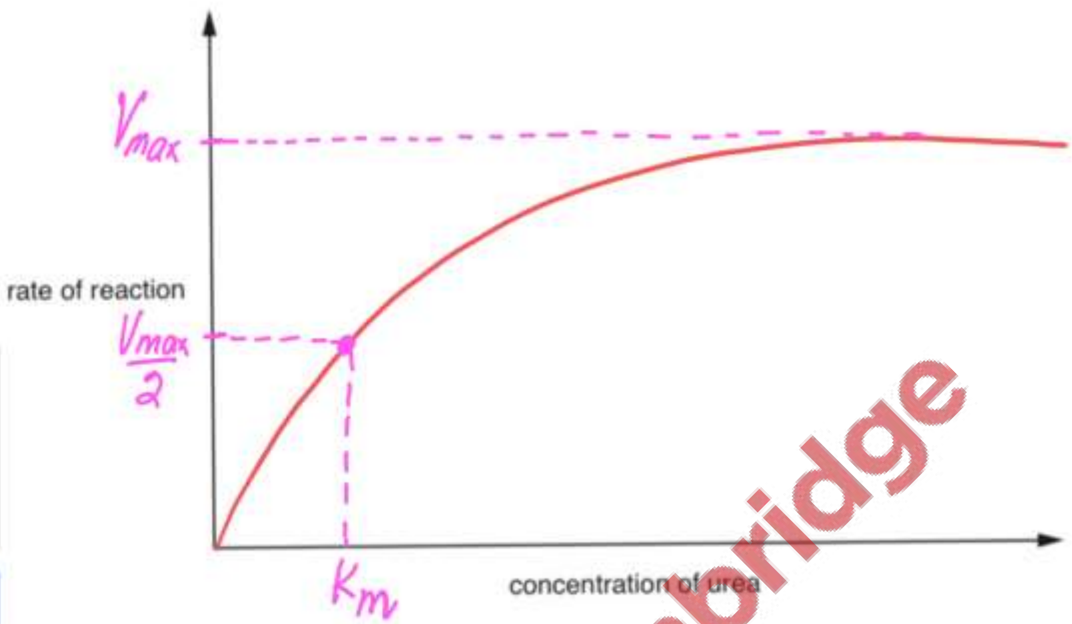
Your method should be set out in a logical way and be detailed enough to let another person follow it. You should **not** include details of how to prepare the urea or urease solutions.

* choose 6 temperature readings:
  10°C, 20°C, 30°C, 40°C, 50°C, 60°C
* use a thermostatically controlled water bath for each temperature
* same volume of urease each time
* same volume of each urea concentration
* same volume of buffer to maintain a constant pH
* Incubate urease and urea concentrations separately
* mix urea and urease and insert the conductivity probe
* take readings from the conductivity meter at the same time for each solution
* test each conc. of urea at each temperature
* minimum of 3 replicates and determine the mean (1)
* low risk experiment (1)

[6]

(c) (i) Use the axes below to sketch a graph to show the effect of substrate concentration on the initial rate of reaction at **one** temperature.

Indicate on your graph how the student could find the K_m of urease at that temperature.



[4]

(ii) Table 1.1 shows the results that the student obtained for K_m at different temperatures.

Table 1.1

temperature	A	B	C	D	E
K_m/arbitrary units	1.8	1.3	0.9	0.8	1.2

State which of these temperatures, **A** to **E**, is closest to the likely optimum temperature of urease.

Explain your answer.

D → temperature with the lowest Km (highest affinity).

[2]

[Total: 19]