DATA ANALYSIS

STATS TEST	USED TO	DEGREES OF FREEDOM	JUSTIFICATION
Student's t-test	Compare the difference between the means of two groups of data	(N1 + N2) - 2 N1 - Sample size Group 1 N2 - Sample size Group 2	Looking for the difference between two means
Chi-squared	Compare the differences between observed and expected results	Number of categories - 1	Using categorical data
Correlation Coefficient (Spearman's Rank)	Test the strength of a correlation between two variables	No degrees of freedom	Looking for correlation / relationship between variables.

Standard Deviation

- shows the spread of data / variation around the mean.
- compared to the range SD reduces the effect of anomalies / outliers
- can be used to determine if the DIFFERENCE between results is SIGNIFICANT (±2 SD contains over 95% of the data)

We use statistical tests to see of the DIFFERENCE (measured change in the dependent variable) is caused by chance OR by the changes we made to the independent variable.

If there is less than 5% ($P \leq 0.05$) probability that the difference in results occurred due to chance then we say the DIFFERENCE is SIGNIFICANT. In other words there is a 95%+ chance that the DIFFERENCE we measured in the dependent variable was caused by the changes we made to the independent variable (and not due to chance)

If you use the term SIGNIFICANT you must also use DIFFERENCE / CORRELATION

✅ The DIFFERENCE is SIGNIFICANT

✅ There is a SIGNIFICANT (positive / negative) CORRELATION

❌ The results are SIGNIFICANT

❌ The practical is SIGNIFICANT

The DIFFERENCE is SIGNIFICANT if

- the test statistic is greater than the critical value at $P = 0.05$ (or lower e.g. $P = 0.01$)
- the ±2 Standard Deviations **do not** overlap

Test statistic: the result of the stats text
Critical value: the value you compare that test statistic against. There is a different critical value for each P-value and each Degree of Freedom (similar to sample size)
P-value (confidence interval / significance level): the **P**robability that the results occurred due to chance

(e) The chi-squared (χ^2) test can be used to analyse the data.

(i) Complete the rows for metaphase and telophase in the table below and calculate the χ^2 value for the data.

The χ^2 value is calculated using the following formula:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Cells	Observed (O)	Expected (E)	(O–E)	$(O–E)^2$	$\frac{(O - E)^2}{E}$
In prophase	85	65	20	400	6.154
In metaphase					
In anaphase	6	7	-1	1	0.143
In telophase					
Total	200	200			

Chi-squared test statistic:
A = 6.92
B = 13.83
C = 18.25

Degrees of freedom =

⚠ **AQA** students **DO NOT** need to calculate any stats test statistic or standard deviation value in the exams

(iii) The student had expected that the results observed for tissue type **W** would not be significantly different from those for tissue type **V**.

Use your calculated value for χ^2 and the information from the χ^2 probability table below to conclude whether or not the results observed for tissue type **W** are significantly different from those for tissue type **V**.

[2]

Degrees of freedom	Probability (p)				
	0.99	**0.95**	**0.05**	**0.01**	**0.001**
1	0.00	0.00	3.84	6.64	10.83
2	0.02	0.10	5.99	9.21	13.82
3	0.11	0.35	7.82	11.35	16.27
4	0.30	0.71	9.49	13.28	18.47
5	0.55	1.15	11.07	15.09	20.52
6	0.84	1.64	12.59	16.81	22.46
7	1.24	2.17	14.07	18.48	24.32

0 4 Scientists investigated the effect of different types of animal farming on the diversity and number of dung beetles. They determined the number of dung beetle species and their total number on intensive (**I**), rough grazing (**R**) and organic (**O**) farms.

Figure 4 and **Figure 5** show some of their results.

Figure 4

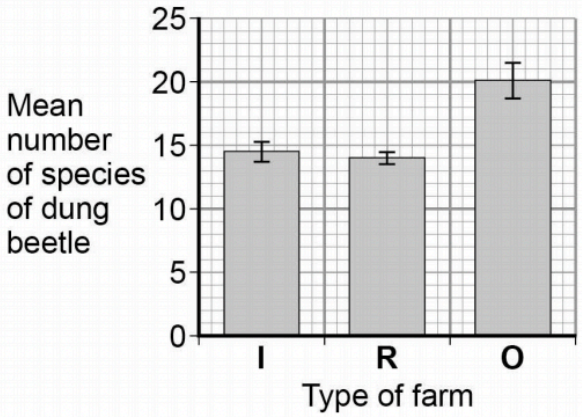
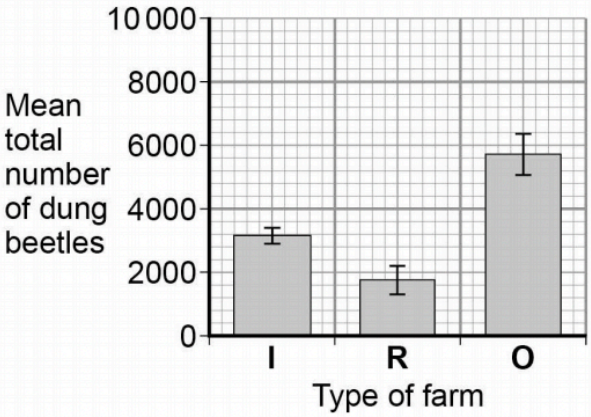


Figure 5



Key: I Standard deviation

0 4 . 3 Explain what the standard deviations suggest about the difference in mean total number of dung beetles between the different types of farm.

[2 marks]

.....

.....

.....

.....

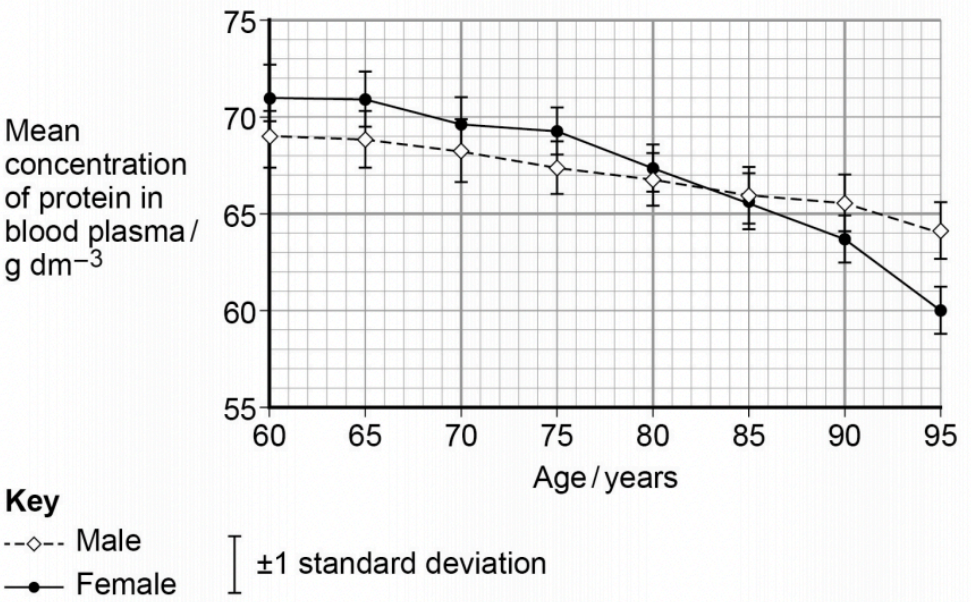
.....

.....

0 4 Scientists investigated how the concentration of protein in blood plasma changes in people between the ages of 60 and 95.

Figure 4 shows the scientists' results. The bars show ±1 standard deviation.

Figure 4



0 4 . 3 What can you conclude from **Figure 4** about the effect of ageing on the mean concentration of protein in the blood plasma in males and females?

[2 marks]

.....

.....

.....

.....

.....

.....

0 5 . 3 **Table 1** shows how the scientists published their results. They calculated mean values and two times the standard deviation (SD) of the mean.

Two standard deviations above and below the mean includes 95.4% of the data.

Table 1

Group of animals	Mean number of animals per m^2 (± 2 × SD)		Mean number of species per m^2 (± 2 × SD)	
	Soil under wheat crop	Soil under grassy strips	Soil under wheat crop	Soil under grassy strips
Beetles	41.2 (± 6.4)	80.1 (± 10.1)	10.0 (± 1.6)	17.3 (± 1.0)
Centipedes	18.4 (± 3.6)	13.5 (± 1.0)	1.8 (± 0.3)	2.1 (± 0.2)
Earthworms	244.5 (± 27.1)	281.2 (± 39.4)	3.8 (± 0.3)	5.1 (± 0.2)
Millipedes	38.4 (± 12.2)	36.2 (± 2.9)	3.5 (± 0.3)	3.2 (± 0.2)
Woodlice	0.0	73.9 (± 8.5)	0.0	2.8 (± 0.2)

0 5 . 4 A summary of this research was published in a farming magazine. The journalist concluded that creating grassy strips around fields had little effect on the diversity of soil animals.

Do you agree with this conclusion?

Use evidence from **Table 1** to justify your answer.

[4 marks]

.....

.....

.....

.....

.....

.....

.....

.....

.....

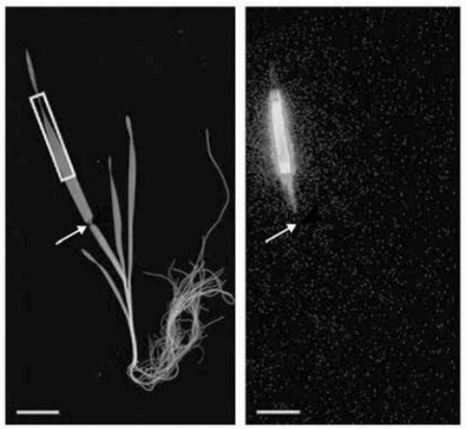
.....

0 8 Scientists investigated the effect of a heat treatment on mass transport in barley plants.

- They applied steam to one short section of a leaf of the heat-treated plants. This area is shown by the arrows in **Figure 9**.
- They did not apply steam to the leaves of control plants.
- They then supplied carbon dioxide containing radioactively-labelled carbon to each plant in the area shown by the rectangular boxes in **Figure 9**.
- After 4 hours, they:
  - found the position of the radioactively-labelled carbon in each plant. These results are shown in **Figure 9**.
  - recorded the water content of the parts of the leaf that were supplied with radioactively-labelled carbon dioxide. These results are shown in **Table 4**.

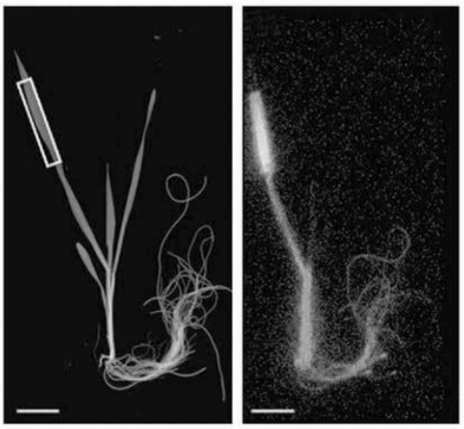
Figure 9

A – Heat-treated Plant



0 hours 4 hours

B – Control Plant, not heat treated



0 hours 4 hours

Table 4

Plant from which the leaf was taken	Water content of leaf / % of maximum (± 2 standard deviations)
Heat-treated Plant A	84.6 (±11.3)
Control Plant, not heat treated B	92.8 (±8.6)

0 8 . 2 The scientists also concluded that this heat treatment did not affect the xylem.

Explain how the results in **Table 4** support this conclusion.

[2 marks]

.....

.....

.....

.....

.....

.....

2,4-D is a selective herbicide that kills some species of plants but not others. 2,4-D disrupts cell-surface membranes but the extent of disruption differs in different species.

Scientists investigated the effect of 2,4-D on wheat plants (a crop) and on wild oat plants (a weed).

They grew plants of both species in glasshouses. They put plants of each species into one of two groups, **W** and **H**, which were treated as follows:

- Group **W** – leaves sprayed with water
- Group **H** – leaves sprayed with a solution of 2,4-D.

After spraying, they cut 40 discs from the leaves of plants in each group and placed them in flasks containing 10 cm^3 de-ionised water. After 5 minutes, they calculated the disruption to cell-surface membranes by measuring the concentration of ions released into the water from the leaf discs.

Their results are shown in **Table 3**.

The lowest significant difference (LSD), is the smallest difference between two means that would be significant at $P \leq 0.05$

Table 3

Group	Treatment	Mean concentration of ions in water / arbitrary units	
		Wheat	Wild oats
W	Water	26	45
H	2,4-D	27	70
Lowest significant difference (LSD)		7	10

0 6 . 2 Evaluate the use of 2,4-D as a herbicide on a wheat crop that contains wild oats as a weed. Use all the information provided.

[4 marks]

.....

.....

.....

.....

.....

.....

.....

.....

.....

Cell walls make up much of the fibre that people eat.

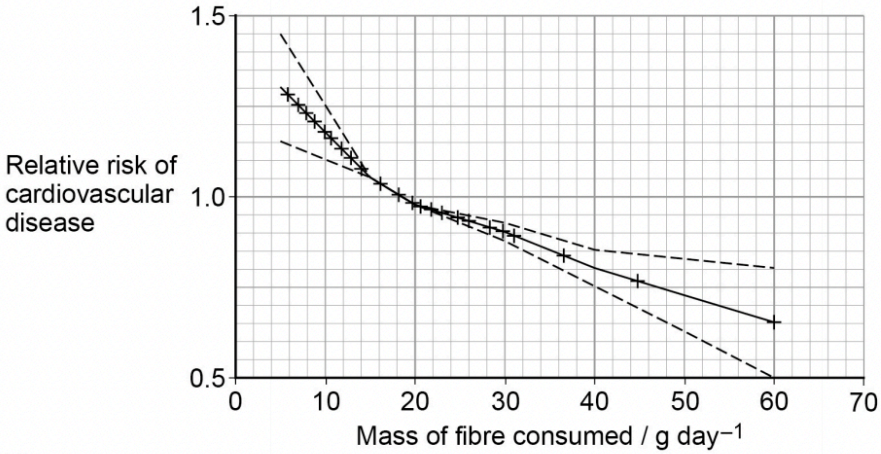
Scientists investigated the relationship between the mass of fibre people ate each day and their risk of cardiovascular disease (CVD).

They gathered data from a large sample of people and used this to calculate a relative risk.

- A relative risk of 1 means there is no difference in risk between the sample and the whole population.
- A relative risk of < 1 means CVD is less likely to occur in the sample than in the whole population.
- A relative risk of > 1 means CVD is more likely to occur in the sample than in the whole population.

Their results are shown in **Figure 3**. A value of ± 2 standard deviations from the mean includes over 95% of the data.

Figure 3



Key

—— Mean relative risk

– – – Line of best fit showing ± 2 standard deviations from the mean

Each '+' plotted point represents 1000 people

0 2 . 2 A student concluded from **Figure 3** that eating an extra 10 g of fibre per day would significantly lower his risk of cardiovascular disease.

Evaluate his conclusion.

[4 marks]

.....

.....

.....

.....

.....

.....

.....

.....

(e) Another group of students repeated this investigation and calculated r_s = 0.589. The critical value of r_s at the 5% level for 9 degrees of freedom is 0.600.

They concluded that their results showed a weak positive correlation between leaf hair density and distance of the tree from the river.

Evaluate the conclusion of this group of students.

.....

.....

..... **[2]**

.....

.....

0 6 . 3 The scientists tested their null hypothesis using the chi-squared statistical test.
After 1 cycle their calculated chi-squared value was 350
The critical value at P=0.05 is 3.841

What does this result suggest about the difference between the observed and expected results and what can the scientists therefore conclude?

[2 marks]

.....

.....

.....

.....

.....

.....

0 4 . 4 The scientists calculated a P value of 0.03 when testing their null hypothesis.

What can you conclude from this result? Explain your answer.

[3 marks]

.....

.....

.....

.....

.....

.....

Scientists investigated the use of a drug called Tenapanor to reduce salt absorption in the gut. Tenapanor inhibits the carrier protein, NHE3.

The scientists fed a diet containing a high concentration of salt to two groups of rats, **A** and **B**.

- The rats in Group **A** were **not** given Tenapanor (0 mg kg^{-1}).
- The rats in Group **B** were given 3 mg kg^{-1} Tenapanor.

One hour after treatment, the scientists removed the gut contents of the rats and immediately weighed them.

Their results are shown in **Table 2**.

Table 2

Concentration of Tenapanor / mg kg^{-1}	**Mean mass of contents of the gut / g**
0	2.0
3	4.1

The scientists carried out a statistical test to see whether the difference in the means was significant. They calculated a P value of less than 0.05.

They concluded that Tenapanor did reduce salt absorption in the gut.

Use all the information provided and your knowledge of water potential to explain how they reached this conclusion.

[4 marks]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Species of butterfly	Mean number of butterflies		P value
	In canopy	In understorey	
Prepona laertes	15	0	< 0.001
Archaeoprepona demophon	14	37	< 0.001
Zaretis itys	25	11	> 0.05
Memphis arachne	89	23	< 0.001
Memphis offa	21	3	< 0.001
Memphis xenocles	32	8	< 0.001

0 3 . 4 The scientists carried out a statistical test to see if the difference in the distribution of each species between the canopy and understorey was due to chance.
The P values obtained are shown in **Table 1**.

Explain what the results of these statistical tests show.

[3 marks]

.....

.....

.....

.....

.....

.....

.....

.....

Actual Chi-squared test statistic = 13.835 or 13.833 or 13.834

Any statement(s) made must be correct for the candidate's responses to (i) and (ii).

two from

1 calculated value is , > / greater than ,7.82 / the critical value at p = 0.05 / the value for (p =) 0.05
or
7.82 / the critical value at p = 0.05 / the value for (p=) 0.05, is , less than / <, 13.835 ✓
2 (difference / deviation) is, significant / not due to chance ✓
3 95% certain that the results are not due to chance
or
difference would only occur by chance 5% of the time ✓
4 (difference / deviation) also significant at p=0.01 value
or
99% certain that the results are not due to chance
or
difference would only occue by chance 1% of the time
or
value is , > / greater than , p = 0.01 / 11.35 or
probability is , < / less than , 0.01
or
probability is between 0.01 and 0.001
or
probability is not significant at p = 0.001 ✓
5 the null hypothesis can be rejected ✓

[2]

ALLOW ecf from candidate's calculated X^2 value in **(i)** using the number of degrees of freedom they stated in **(ii).**

For incorrect X^2 and degrees of freedom values, apply mark points 1 to 5 to correspond to their results.

Degrees of freedom	Probability (p)				
	0.99	0.95	0.05	0.01	0.001
1	0.00	0.00	3.84	6.64	10.83
2	0.02	0.10	5.99	9.21	13.82
3	0.11	0.35	7.82	11.35	16.27
4	0.30	0.71	9.49	13.28	18.47
5	0.55	1.15	11.07	15.09	20.52
6	0.84	1.64	12.59	16.81	22.46
7	1.24	2.17	14.07	18.48	24.32

0 4 . 3 Explain what the standard deviations suggest about the difference in mean total number of dung beetles between the different types of farm.

[2 marks]

1. No overlap in standard deviations;
2. (Difference in mean total) significant/is not due to chance/is real;

Accept: no overlap in error bars.

0 4 . 3 What can you conclude from **Figure 4** about the effect of ageing on the mean concentration of protein in the blood plasma in males and females?

[2 marks]

1. Protein content decreases with age and decreases more in females;
2. Difference (between sexes) only significant at 95 years because SDs do not overlap;
OR
Differences not significant because 2 x SD would overlap;

0 5 . 4 A summary of this research was published in a farming magazine. The journalist concluded that creating grassy strips around fields had little effect on the diversity of soil animals.

Do you agree with this conclusion?

Use evidence from **Table 1** to justify your answer.

[4 marks]

Principle:
1. Overlap of 2 × SD shows probability of differences (in means) being due to chance is greater than 0.95;
Agree:
2. No difference in number of earthworms and millipedes (per m2);
3. No difference in number of species of centipedes or millipedes;
Disagree:
4. More beetles and woodlice in grassy strips;
5. More species of beetles, earthworms, woodlice in grassy strips;

0 8 . 2 The scientists also concluded that this heat treatment did not affect the xylem.

Explain how the results in **Table 4** support this conclusion.

[2 marks]

1. (The water content of the leaves was) not different because (means ± 2) standard deviations overlap;
2. Water is (therefore) still being transported in the xylem (to the leaf)
OR
Movement in xylem is passive so unaffected by heat treatment;

1. For 'not different' accept 'difference is not significant' or 'difference due to chance'.

0 6 . 2 Evaluate the use of 2,4-D as a herbicide on a wheat crop that contains wild oats as a weed. Use all the information provided.

[4 marks]

1. 2,4-D causes an increase in release of ions from wild oat cells and 2,4-D does not affect/has little effect on the release of ions from wheat cells;
2. (For wheat) Difference is less than LSD/7 so difference is not significant;
OR
(For wild oats) Difference is more than LSD/10 so difference is significant;
3. Loss of ions from cells (likely to) lead to cell/plant death/damage;
OR
Disruption of cell membrane (likely to) lead to cell/plant death/damage;
4. No evidence here about death of plants as a result of this ion loss;
5. No evidence here of other ecological/environmental impact;

1. Accept reference to 'concentration of ions in water' or 'disruption of the cell membranes' in place of 'release of ions'
1. Accept 'difference in release of ions from wild oats is 25 and difference in release of ions from wheat is 1'
2. Accept
'(For wheat) difference is less than LSD so greater than 5% probability that difference is due to chance'
OR
'(For oats) difference is more than LSD so less than 5% probability that difference is due to chance'
5. Accept 'development of resistance'

outside the box

0 2 . 2 A student concluded from **Figure 3** that eating an extra 10 g of fibre per day would significantly lower his risk of cardiovascular disease.

Evaluate his conclusion.

[4 marks]

1. Negative correlation (between fibre eaten per day and risk of cardiovascular disease);
2. Original/current fibre intake (of student) not known;
3. (Idea of) significance linked to (2x) standard deviation overlap (at 10 g day^{-1} change);

4. If current intake between 5 and 30 (g day-1) then (eating 10g more results in a significant) decrease in risk

OR

If current intake between 30 and 50 (g day-1) then (eating 10g more results in) no significant decrease in risk;

5. Correlation does not mean causation

OR

Another named factor may be involved;

6. Little evidence/data for higher mass of fibre per day;
7. Large (2x) standard deviation at high/low mass of fibre makes (mean) less precise

OR

Large (2x) standard deviation at high/low amounts of fibre means there is a greater uncertainty;

8. No statistical test (to show if differences are significant);

1. Accept positive correlation with reduced risk
2. Accept 'it depends on original/current fibre intake'.
3. This is for the correct concept, ignore stated values.
3. Ignore reference to probability and chance.
4. Accept stated values between 5 and 30 for (significant) decrease in risk.
4. Accept stated values between 30 and 50 for no significant decrease in risk.
4. Ignore stated values less than 5 or more than 50.
5. Examples of named factors - smoking, exercise, age, sex, genes, other aspects of diet.
7. For 'precise' accept reliable or description of precise/reliable.

(e) Another group of students repeated this investigation and calculated r_s = 0.589. The critical value of r_s at the 5% level for 9 degrees of freedom is 0.600.

They concluded that their results showed a weak positive correlation between leaf hair density and distance of the tree from the river.

Evaluate the conclusion of this group of students.

their conclusion is incorrect ✓
reject (the student's), hypothesis / H_1
there is no, relationship / correlation, (between leaf hair density and distance from river) / the pattern seen is due to chance ✓

[2]

ORA accept the null hypothesis / H_0

0 6 . 3 The scientists tested their null hypothesis using the chi-squared statistical test.
After 1 cycle their calculated chi-squared value was 350
The critical value at P=0.05 is 3.841

What does this result suggest about the difference between the observed and expected results and what can the scientists therefore conclude?

[2 marks]

1. There is a less than 0.05/5% probability that the difference(s) (between observed and expected) occurred by chance;
2. Calculated value is greater than critical value so the null hypothesis can be rejected;
3. (The scientists can conclude that) the proportion of plants that produce 2n gametes does change from one breeding cycle to the next;

1. Reject 'results (without reference to difference) occurring by chance'. Overall max 1 with this statement.
1. Accept 'there is a greater than 0.95/95% probability that the difference did **not** occur by chance'.
1. and 2. Ignore 'difference is significant'
2. Do not accept 'P value' for 'critical value'.

0 4 . 4 The scientists calculated a P value of 0.03 when testing their null hypothesis.

What can you conclude from this result? Explain your answer.

[3 marks]

1. Probability that difference (in frequency of births above 4500 g) is due to chance is **less than** 0.05
OR
Probability that difference (in frequency of births above 4500 g) is due to chance **is** 0.03;

2. Reject null hypothesis;
3. Presence of *KIR2DS1*/allele does (significantly) affect the frequency of high birth mass;

Scientists investigated the use of a drug called Tenapanor to reduce salt absorption in the gut. Tenapanor inhibits the carrier protein, NHE3.

The scientists fed a diet containing a high concentration of salt to two groups of rats, **A** and **B**.

- The rats in Group **A** were **not** given Tenapanor (0 mg kg^{-1}).
- The rats in Group **B** were given 3 mg kg^{-1} Tenapanor.

One hour after treatment, the scientists removed the gut contents of the rats and immediately weighed them.

Their results are shown in **Table 2**.

Table 2

Concentration of Tenapanor / mg kg^{-1}	Mean mass of contents of the gut / g
0	2.0
3	4.1

The scientists carried out a statistical test to see whether the difference in the means was significant. They calculated a P value of less than 0.05.

They concluded that Tenapanor did reduce salt absorption in the gut.

Use all the information provided and your knowledge of water potential to explain how they reached this conclusion.

[4 marks]

1. Tenapanor/(Group)B/drug causes a significant increase;
OR
There is a significant difference with Tenapanor/drug/between A and B;
2. There is a less than 0.05 probability that the difference is due to chance;
3. (More salt in gut) reduces water potential in gut (contents);
4. (so) less water absorbed out of gut (contents) by osmosis
OR
Less water absorbed into cells by osmosis
OR
Water moves into the gut (contents) by osmosis.
OR
(so) water moves out of cells by osmosis;

1. and 2. Reject references to 'results' being significant/due to chance once only.
2. Do not credit suggestion that probability is 0.05% or 5.
2. Accept 'There is a greater than 0.95/95% probability that any difference between observed and expected is not due to chance

0 3 . 4 The scientists carried out a statistical test to see if the difference in the distribution of each species between the canopy and understorey was due to chance. The P values obtained are shown in **Table 1**.

Explain what the results of these statistical tests show.

[3 marks]

1. For Zaretis itys, difference in distribution is probably due to chance / probability of being due to chance is more than 5%;
2. For all species other than Zaretis itys, difference in distribution is (highly) unlikely to be due to chance;
3. Because P < 0.001 which is highly significant/is much lower than 5%;