**Statistics in AS Biology**

**In AS exams, students could be expected to:**

Formulate a null hypothesis: for the experiments they perform during their class work **and** when given appropriate information, for experiments carried out by others.

**Devise** and **justify** an appropriate table in which to record their raw data.

**Devise** and **justify** an appropriate way to represent their processed data graphically.

**Select** and **justify** the selection of an appropriate statistical test for data they will subsequently collect themselves or data that might be collected by others. The statistical tests are restricted to:

• chi-squared test when the data are categorical

• the Student’s t test when comparing the mean values of two sets of data

• a correlation coefficient when examining an association between two sets of data.

**Interpret a given probability value** in terms of the **probability** of the difference between observed data and expected data (chi-squared test), the difference between the means of two samples (Student’s t test) or a correlation between two variables (correlation coefficient) **being due to chance.**

This is most easily done by using the following framework: “…… this shows there is a greater/less than **5% probability** that the difference/correlation between ……… is **due to chance**.Therefore the null hypothesis should be accepted/can be rejected.”

In written examinations, students might be asked to perform simple calculations such as finding a mean value.

Students will **not** be asked to perform a calculation using a statistical test (or to calculate the standard deviation of a mean). We would expect students will perform such calculations during their class work.

**Statistical tests and tables of critical values**

 **Tables of critical values**

A table of critical values is provided with each statistical test. If your calculated test statistic is greater than, or equal to, the critical value, then the result of your statistical test is significant. This means that your null hypothesis should be rejected.

**Spearman rank correlation test**

Use this test when you wish to find out if there is a significant **association** between **two** sets of measurements from the **same sample** and you have between 5 and 30 **pairs** of measurements.

Record the data as values of X and Y

Convert these values to rank orders, 1 for largest, 2 for second largest, etc.
If two or more values are the same use the average rank for each.

Now calculate the value of the Spearman rank correlation, rs, from the equation



Where N is the number of pairs of items in the sample.

D is the difference between each pair (X-Y) of ranked measurements.

 **A table showing the critical values of rs for different numbers of paired values.**



The **Student t** **test**

Use this test when you wish to find out if there is a **significant difference** between **two means** when the data are normally distributed and the **sample size** is less than 25.

 t can be calculated from the formula



Where:

x1 = mean of first sample

x2 = mean of second sample

 s1 = standard deviation of first sample

s2 = standard deviation of second sample

n1 = number of measurements in first sample

n2 = number of measurements in second sample

**A table showing the critical values of t for different degrees of freedom.**



 The number of degrees of freedom = (n1 + n2) – 2

**The chi-squared (χ2) test**

Use this test when the measurements relate to the number of individuals in particular categories; the observed number can be compared with an expected number which is calculated from a theory, as in the case of genetics experiments.

The chi-square (χ2) test is based on calculating the value of χ2 from the equation



Where O represents the observed results and E represents the results we expect.

**A table showing the critical values of χ2 for different degrees of freedom.**

****

Degrees of freedom can be calculated in different ways, depending on the data:

the simplest is: (number of categories -1)

A better alternative, if there are >1 row *and* column, is: (no. rows-1) x (no. columns-1)

**Students’ Statistics Sheet**

****