Formulae Sheet

Notes on the use of skills, techniques and statistics in Edexcel GCE Geography specification A

The specification requires candidates to be able to demonstrate an ability within a geographical context to:

- Collect, record and interpret a variety of evidence from a range of first-hand and secondary sources;
- Select from a range of enquiry methodologies and apply them appropriately;
- Organise and present information, ideas, descriptions and arguments clearly and logically, taking into account their use of grammar, punctuation and spelling;
- Demonstrate skills of analysis, synthesis and critical evaluation;
- Use quantitative and qualitative investigative techniques.

These skills will be assessed through all the units. Skills and techniques which are specific to each units are identified in the units description.

The skills and techniques expected of candidates are listed on pages 16 and 17 in the specification. If a statistical technique is used in an examination question, then it will be written as in the table below. **When used in the questions, the formula will always be given.**

When using formulae in an examination, candidates will receive some credit for correctly substituting the values in the formula and some for the calculation of the correct answer. Careless errors will not be double-penalised, ie should the substitution into the formula be incorrect, provided the calculation correctly handles this incorrect substitution, marks will be awarded for the calculation.

Candidates should be able to interpret the result and comment not only on its significance, but also the general application of the technique.

Candidates should:

a. Understand the need for the classification of data and be able to present data as histograms or frequency diagrams;

b. Know the main sampling methods used in geography and the importance of sampling in gaining information about a population, and appreciate the meaning of sample error;

c. Possess skills in graphicacy in order to represent and interpret the data in the form of line graphs, bar and compound graphs, circular graphs and triangular graphs;

d. Be able to compute means and medians, and interpret information about dispersion (variation), including the range and quartiles;

e. Understand simple correlation, including the calculation of Spearman rank coefficient and chi square, and be able to apply checks of significance from a table of values;

f. Be able to sketch a best fit line form an array of points on a scatter diagram and identify positive and negative anomalies.

Statistic	Formula	Note of computational form
1. Statistical techniques used across all units		
Sample arithmetic mean	$\bar{x} = \frac{x}{n}$	
Median	Location = $\frac{n+1}{2}$	when the values are arranged in rank order
Range	Range = highest value − lowest value	
Interquartile range	Interquartile range = $Q_3 - Q_1$	Q_3 is the upper quartile Q_1 is the lower quartile
Spearman rank correlation coefficient	$r_s = 1 - \frac{6 \quad d^2}{(n^3 - n)}$	d is the difference in the ranking for each item (**Note: when tied ranks occur, the ranks are to be averaged**)
Chi square	$\chi^2 = \frac{(O-E)^2}{E}$	O is the observed frequency E is the expected frequency
2. Statistical techniques introduced in specific units. Techniques introduced in AS units may also be used in A2 units. Techniques introduced in an A2 unit will be used in that unit only		
Area nearest neighbour (unit 2.1)	$R_n = 2\bar{d}\sqrt{\frac{n}{a}}$	$\bar{d}$ is the mean of observed distance between nearest neighbours n is the number of places a is the area
Location quotient (unit 5.1)	$LQ = \frac{P_{ij}}{P_{ik}}$	P_{ij} is the proportion of the population with the characteristic i in area j P_{ik} is the proportion of the population with the characteristic i in area k where area j is a region of area k
Hydraulic radius (unit 1.2)	$R = \frac{A}{p}$	A is the cross sectional area of the channel P is the wetted perimeter
Dependency ratio (unit 2.3)	$\frac{\text{children (0 - 14)} + \text{elderly (over 65)}}{\text{all population (15 - 65)}} \times 100$	
Gravity model (unit 2.3)	$I_{ij} = \frac{P_i \times P_j}{(d_{ij})^k}$	I_{ij} flow from place i from place j

		P_j population (or another measure) d_{ij} is the distance k is a constant
Breakpoint (unit 2.2)	$1 + \frac{d_{ij}}{P_i \div P_j}$	d_{jk} is the distance P is the population

1.)**Testing for an association between two sets of data**
If the data is ordinal or interval data, the Spearman Rank test should be used:

$$r_s = 1 - \frac{6 \quad d^2}{(n^3 - n)}$$

where d = the difference in rank (when tied ranks occur, the ranks are averaged)

If the data is of the categorial type, the chi squared test should be used:

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

where: O = observed frequency
E = expected frequency

Both of these calculations seek to prove a Null Hypothesis: this is a statement that there is no relationship between the two variables. Tables are used to interpret the significance of the result and to see with what confidence the Null Hypothesis can be rejected. The appropriate table(s) will be provide for use by candidates in examinations where these tests are applied.

2.) Measuring dispersions
Nearest neighbour produces an index which provides a test for the 'non-randomness' of a certain distribution. The index ranges from 0 (ie points completely clustered) to 2.15 (ie a totally uniform distribution, where points are spread as far apart from each other as is possible).

The formula used is:

$$R_n = 2\overline{d}\sqrt{\frac{n}{a}}$$

Where: n = total number of points in survey
$\overline{d}$ = mean of the distance measured between each point and its nearest neighbour
a = area of survey