Department for Education

Statistics

Draft GCSE subject content

September 2015

Contents

The content for statistics GCSE

Introduction

1. The GCSE subject content sets out the knowledge, understanding and skills common to all specifications in GCSE statistics. Specifications in GCSE statistics must require students to apply the statistical techniques listed in paragraph 8 through the practical investigation of real world data collected from authentic contexts within the framework of the statistical enquiry cycle, as detailed in appendix 3.

Aims and objectives

2. GCSE specifications in statistics must encourage students to develop statistical literacy through:

- using statistical techniques in a variety of authentic investigations, using real world data (both primary and secondary as appropriate) in contexts such as, but not limited to, populations, climate, sales etc.
- identifying trends through carrying out appropriate calculations and data visualization techniques
- the application of statistical techniques across the curriculum, in subjects such as the sciences, social sciences, computing, geography, business and economics, and outside of the classroom in the world in general
- critically evaluating their own data, calculations and evaluations, and of a range of published data, calculations and evaluations that would be commonly encountered in their studies and in everyday life
- gaining direct experience of using some of the technologies that have enabled the collection, visualisation and analysis of large quantities of data to inform decision-making processes in public, commercial and academic sectors
- applying appropriate mathematical and statistical formulae, as set out in Appendix 1, and building upon the numerical skills developed in GCSE mathematics, as listed in appendix 2

Subject content

Overview

3. The subject criteria set out the full range of content for GCSE specifications in statistics. Awarding organisations may, however, use any flexibility to increase depth, breadth or context within the specified topics or to consolidate teaching of the subject content.

4. Specifications should ensure students develop the confidence and competence with statistical techniques that enable them to apply those techniques flexibly to solve statistical problems.

5. The expectation of the GCSE statistics content is that:

- all students will develop confidence and competence with the content identified by standard type
- all students will be assessed on the content identified by the standard and the underlined type in curly brackets, underline type in curly brackets will be included in both foundation and higher papers; more highly attaining students will develop confidence and competence with all of this content
- only the more highly attaining students will be assessed on the content identified by bold type and in square brackets; the highest attaining students will develop confidence and competence with this content

6. The distinction between standard, underlined type in curly brackets and bold type in square brackets applies to the content statements only, not to the assessment objectives or to the statistical enquiry cycle and numerical skills in the appendices.

7. GCSE specifications in statistics must require students to demonstrate an understanding of the statistical enquiry cycle, as detailed in Appendix 3, relating to the research methodologies used in statistical analysis. In carrying out practical statistical investigations, students must utilise their knowledge of the subject content to appreciate the role statistics plays in the decision making processes used in the world around them by using real data taken from authentic contexts. Students must recognise that different approaches, including the use of technology, may be appropriate at each stage of the statistical enquiry cycle, and that statistical conclusions are developed through an iterative process of retesting and refinement.

Detailed subject content

8. GCSE specifications in statistics must require students to:

A	**Understand the importance of the careful planning of a clear strategy for collecting, recording and processing data in order to address an identified question or hypothesis**
A1	Know that a hypothesis can only be tested through the appropriate collection and analysis of data. Terminology such as null hypothesis will not be required
A2	Know the constraints that may be faced in designing an investigation to test an hypothesis: these may include factors such as time, costs, ethical issues, confidentiality, and convenience, etc.
A3	Determine proactive strategies to mitigate issues that might arise during the statistical enquiry process. For example, dealing with difficulties in identifying the population, non-response issues or unexpected outcomes

B	**Recognise the opportunities, constraints and implications for subsequent mathematical analysis involved in obtaining appropriate data through careful design of primary data collection techniques or through the use of reference sources for secondary data to ensure unbiased research**
B1	Know and apply terms used to describe different types of data that can be collected for statistical analysis: raw data, {quantitative, qualitative}, categorical, ordinal, discrete, continuous, ungrouped, grouped, {bivariate} and **[multivariate]** Know the advantages and implications of merging data into more general categories, and of grouping numerical data into class intervals {Know and apply the terms explanatory or independent variables and response or dependent variables}
B2	Know the difference between primary and secondary data Know that data can be collected from different sources: experimental (laboratory, field and natural), simulation, questionnaires, observation, reference, census, population and sampling and that sources of secondary data should be acknowledged {Know the importance of reliability and validity with regards to collected data} {Determine factors that may lead to bias, including issues of sensitivity of the content matter, and know how to minimise data distortion, including level of control}
B3	Know the difference between population, sample frame and sample Know that "population" can have different meanings within a stated context Know reasons for employing judgement or opportunity (convenience) sampling, and the associated risks of bias when these techniques are used Use appropriate sampling techniques in the context of the problem to avoid bias: random, systematic, and {quota} Know the key features of a simple random sample {and apply different techniques, both physical and electronic, to select random members from a population: e.g. dice, cards, random number lists, calculator functions and software functions} {Use stratification and know when this is appropriate before sampling takes place}
B4	Know the key features to be considered when planning data collection: leading questions, avoiding biased sources, time factors, open/closed questions, different types of interview technique
B5	Know and apply techniques to deal with problems that may arise with collected data: for example, missing data, incorrect formats, non-responses, incomplete responses, etc. Know why data may need to be "cleaned" before further processing, including issues that arise on spreadsheets and apply techniques to clean data in context {Know the importance of identifying and controlling extraneous variables} **[and the use of control groups]**

C	**Generate data visualisation through the use of technology and by manual construction and understand the mathematics required to derive these visualisations**
C1	Represent data sets pictorially using calculated key values as necessary, and interpret and compare data sets displayed pictorially: tabulation, tally, pictogram, {pie chart, stem and leaf diagram, Venn diagram} Interpret and compare data sets displayed pictorially: {population pyramid, choropleth map}, **[comparative pie chart, comparative 2D representations, comparative 3D representations]**
C2	Represent data sets graphically using calculated key values as necessary, and interpret and compare data sets displayed graphically: bar charts, line charts, time series, scatter charts, {bar line charts, frequency polygons, cumulative frequency (discrete and grouped) charts, histograms (equal width), and box plots} **[Calculate and use frequency density to draw histograms (unequal width), and interpret and compare data sets displayed in histograms (unequal width)]**
C3	Represent data by selecting appropriate format and produce accurate visualisation of data Recognise where errors in construction lead to graphical misrepresentation, including but not limited to incorrect scales, truncated axis, distorted sizing **[or the misuse of formula when calculating the frequency densities of histograms]**
C4	Extract and calculate corresponding values in order to compare data sets that have been presented in different formats and be able to present the same information in multiple formats Select {and justify} appropriate form of representation {with regard to the nature of data}

D	**Calculate statistical measures to compare data**
D1	Calculate averages for discrete and grouped data: mode, median, arithmetic mean, {weighted mean}, **[geometric mean, mean seasonal variation]** {Justify the rationale for selecting appropriate types of average in context}
D2	Determine skewness from data by inspection {and by calculation. Formula to be given when required}
D3	Calculate different measures of spread using IT and manually: range, quartiles, {interquartile range, percentiles}, **[interpercentile range and interdecile range],** {variance and standard deviation} Identify outliers by inspection {and using appropriate calculations}
D4	Identify trends in data through inspection and {by calculation of moving averages}
D5	Determine line of best fit by eye, {by drawing through a calculated double mean point $(\bar{x}, \bar{y})$} **[and by calculating the equation of the regression line]**
D6	**[Apply appropriate formulae to calculate correlation: Spearman's rank**

	correlation coefficient formula, Pearson's product moment correlation coefficient formula. Values found using calculator functions will be permissible]
D7	Use collected data to calculate estimates of probabilities
D8	**[Calculate the probability and expected frequency of outcomes in a binomial distribution. Values found using calculator functions will be permissible]**
D9	**[Calculate the probability and expected frequency of outcomes in a Normal distribution. Values found using calculator functions will be permissible]**

E	**Use visualisation and calculation to interpret results with reference to the context of the problem, and to evaluate the validity and reliability of statistical findings**
E1	Compare the probability of different possible outcomes using the 0-1 or 0-100% scale Use probability values to calculate expected frequency of a specified characteristic within a sample or population {Use collected data and calculated probabilities to determine and interpret relative risks and absolute risks, and express in terms of expected frequencies in groups}
E2	Compare experimental data with theoretical predictions to identify possible bias within the experimental design Recognise that experimental probability will tend towards theoretical probability as the number of trials increases when all variables are random {Use two-way tables, sample space diagrams, tree diagrams and Venn diagrams to represent all the different outcomes possible for at most three events}
E3	Compare different data sets using appropriate calculated or given measure of central tendency: mode, modal group, median and mean Compare different data sets using appropriate calculated or given measure of spread: range, {interquartile range, percentiles, variance and standard deviation} {Use calculated or given mean and standard deviation or median and interquartile range} **[or interpercentile range or interdecile range]** {to compare data samples and to compare a sample with population data}
E4	{Know and apply the formal notation for independent events {$P(A \cap B) = P(A)P(B)$} {$P(A\|B) = P(A)$ and $P(B\|A) = P(B)$}}
E5	Interpret a distribution of data in terms of skewness identified from inspection {or calculation} {Comment on outliers with reference to the original data}
E6	{Interpret seasonal and cyclic trends in context} **[Use such trends to make predictions]**

E7	Interpret data related to rates of change over time (including, but not limited to, births, deaths, house prices, and unemployment) when given in graphical form. {Calculate and interpret rates of change over time from tables using context specific formula} {Use different types of index and weighted index numbers in context, including but not limited to retail price index, consumer price index and gross domestic product}
E8	Know and apply vocabulary of correlation: positive, negative, zero, causation, association, interpolation and extrapolation Make comparisons of correlation by inspection: strong or weak {Know that correlation does not necessarily imply causation} **[Know that there are multiple factors that may interact]**
E9	{Interpret} **[calculated or]** {given Spearman's rank correlation coefficient or Pearson's product moment correlation coefficient in the context of the problem}
E10	{Know **[and interpret]** {the characteristics of a binomial distribution} {Comment on the differences between experimental and theoretical values in terms of possible bias}. Formal tests of significance will not be required
E11	{Know **[and interpret]** {the characteristics of a Normal distribution} {Know that, for a Normal distribution, values more than three standard deviations from the mean are very unusual; know that approximately 95% of the data lie within two standard deviations of the mean and that 68% (just over two thirds) lie within one standard deviation of the mean} {Use action and warning lines in quality assurance sampling applications} **[Use calculated or given means and standard deviation to standardise and interpret data collected in two comparable samples. Formulae will be given]**
E12	{Use calculated or given summary statistical data to make estimates of population characteristics. Use samples to estimate population mean. Use sample data to predict population proportions} **[Use Petersen capture/recapture method to calculate an estimate of the size of a population]**
E13	Know that sample size has an impact on reliability and replication **[Know that a set of sample means are more closely distributed than individual values from the same population]**

Appendix 1 - formulae

Foundation tier

GCSE Maths formulae that may be required:

Specification reference	Formula required
B3	In order to carry out stratification, calculating the percentage or proportion of an amount: $\frac{x}{100} \times$ amount
C1	Calculating the angle for a sector in a pie chart: $\frac{x}{total} \times 360$
C2	Frequency density for a histogram: *frequency density* = $\frac{frequency}{classwidth}$
D1	Calculation of arithmetic mean: $\bar{x} = \frac{\sum fx}{\sum f}$
D3	Range = highest value – lowest value
D3	Interquartile range (IQR) = upper quartile – lower quartile
D7	$Probability = \frac{Number\ of\ favourable\ outcomes}{Total\ number\ of\ outcomes}$

Statistical formulae that students are expected to recall:

Specification reference	Formula required
D1	Calculation of weighted mean: weighted mean = $\frac{\sum(value \times weight)}{\sum weights}$
D3	Identification of an outlier: Small outlier is < LQ – 1.5IQR Large outlier is > UQ + 1.5IQR Outlier is also outside $\mu \pm 3\sigma$
D4	Moving averages (m.a.) e.g. 5 point moving average:

	m.a. = $\frac{x_1+x_2+x_3+x_4+x_5}{5}$
D5	Double mean point: calculate arithmetic mean as in D1
E4	Formulae for independent events: P(A∩B) = P(A)P(B) P(A I B) = P(A) and P(B I A) = P(B)
E7	Index number = $\frac{\text{current value of item}}{\text{value in base year}} \times 100$
E7	Weighted index number = $\frac{\sum(index \times weight)}{\sum weights}$

Complex GCSE statistics formulae to be given in assessment:

Specification reference	Formula required
D2	Skew = $\frac{3(mean-median)}{\text{standard deviation}}$
D3	Population variance = $\frac{1}{N}\sum(x-\mu)^2$ Population standard deviation = $\sqrt{\frac{1}{N}\sum(x-\mu)^2}$
D3	Sample variance = $\frac{1}{N-1}\sum(x-\overline{x})^2$ Sample standard deviation = $\sqrt{\frac{1}{N-1}\sum(x-\overline{x})^2}$
E7	Rates of change (e.g. birth rate = $\frac{\text{number of births}}{\text{total population}} \times 1000$)

Higher tier

GCSE Maths formulae that may be required:

Specification reference	Formula required
B3	In order to carry out stratification, calculating the percentage or proportion of an amount: $\frac{x}{100} \times$ amount

C1	Calculating the angle for a sector in a pie chart: $\frac{x}{total} \times 360$
C2	Frequency density for a histogram: *frequency density* = $\frac{frequency}{classwidth}$
D1	Calculation of arithmetic mean: $\bar{x} = \frac{\sum fx}{\sum f}$
D3	Range = highest value – lowest value
D3	Interquartile range (IQR) = upper quartile – lower quartile
D7	$Probability = \frac{Number\ of\ favourable\ outcomes}{Total\ number\ of\ outcomes}$

Statistical formulae that students are expected to recall:

Specification reference	Formula required
C1	Calculating the radius of a circle for a proportional pie chart: $r = \sqrt{\frac{newtotal \times oldradius^2}{oldtotal}}$
D1	Calculation of weighted mean: weighted mean = $\frac{\sum(value \times weight)}{\sum weights}$
D1	Geometric mean = $\sqrt[n]{value_1 \times value_2 \times \ldots \times value_n}$
D3	Interpercentile range and interdecile range as appropriate: percentile 1 – percentile 2 (calculation of a percentage needed as in B3 above)
D3	Identification of an outlier: Small outlier is < LQ – 1.5IQR Large outlier is > UQ + 1.5IQR Outlier is also outside $\mu \pm 3\sigma$
D4	Moving averages (m.a) e.g. 5 point moving average: m.a. = $\frac{x_1 + x_2 + x_3 + x_4 + x_5}{5}$
D5	Double mean point: calculate arithmetic mean as in D1
D5	Equation of regression line $y = a + bx$ or $y = mx + c$ where gradient (b or m) is:

	$\frac{y_2 - y_1}{x_2 - x_1}$
E4	Formulae for independent events: P(A∩B) = P(A)P(B) P(A I B) = P(A) and P(B I A) = P(B)
E7	Index number = current value of item × 100 value in base year
E7	Weighted index number = $\frac{\sum(index \times weight)}{\sum weights}$
E13	Peterson capture - recapture formula: Number in population = sample size 1 × sample size 2 number marked in sample 2

Complex GCSE statistics formulae to be given in assessment.

Specification reference	Formula required
D2	Skew = $\frac{3(mean - median)}{\text{standard deviation}}$
D3	Population variance = $\frac{1}{N}\sum(x-\mu)^2$ Population standard deviation = $\sqrt{\frac{1}{N}\sum(x-\mu)^2}$
D3	Sample variance = $\frac{1}{N-1}\sum(x-\overline{x})^2$ Sample standard deviation = $\sqrt{\frac{1}{N-1}\sum(x-\overline{x})^2}$
D6	Spearman's rank correlation coefficient $r_s = 1 - \frac{6\sum d_i^{\,2}}{n(n^2-1)}$
D6	Pearson's product moment correlation coefficient $r = \frac{S_{xy}}{\sqrt{S_{xx} \times S_{yy}}} = \frac{\sum(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\{\sum(x_i-\overline{x})^2\}\{\sum(y_i-\overline{y})^2\}}} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sqrt{\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)\left(\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right)}}$

E7	Rates of change (e.g. birth rate = $\frac{\text{number of births}}{\text{total population}} \times 1000$)
There will be an expectation that learners would access the following formulae using their calculator:	
D8	Binomial probability calculations: $P(X = x) \binom{n}{x} p^x (1-p)^{n-x}$
D9	Normal probability calculations: $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
E12	Standardise a normal distribution $Z = \frac{X - \mu}{\sigma}$

Appendix 2 - numerical skills required for GCSE statistics

Integers, fractions, decimals and percentages

- work interchangeably with terminating decimals and their corresponding fractions (such as 3.5 and $\frac{7}{2}$ or 0.375 or $\frac{3}{8}$). Recognise that some fractions can be written as recurring decimals
- identify and work with fractions in ratio problems
- interpret fractions and percentages as operators

Structure and calculation

- order positive integers, decimals and fractions
- understand and use the symbols =, ≠, <, >, ≤, ≥
- apply the four operations to integers, decimals and simple fractions (proper and improper), and mixed numbers
- understand and use place value (e.g. when working with very large or very small numbers, and when calculating with decimals)
- understand and use standard form
- recognise and use relationships between operations, including inverse operations e.g. cancellation to simplify calculations and expressions; use conventional notation for priority of operations, including brackets, powers, roots and reciprocals
- substitute numerical values into formulae and expressions, including scientific formulae
- understand and use standard mathematical formulae; rearrange formulae to change the subject
- work with coordinates on Cartesian grid

Measures and accuracy

- use standard units of mass, length, time, money and other measures (including standard compound measures) using decimal quantities where appropriate
- estimate answers; check calculations using approximation and estimation, including answers obtained using technology
- use compound units such as speed, rates of pay, unit pricing
- round numbers and measures to an appropriate degree of accuracy (e.g. to a specified number of decimal places or significant figures); use inequality notation to specify simple error intervals due to truncation or rounding

Ratio, proportion and rates of change

- express one quantity as a fraction of another, where the fraction is less than 1 or greater than 1
- use ratio notation, including reduction to simplest form
- divide a given quantity into two parts in a given part:part or part:whole ratio; express the division of a quantity into two parts as a ratio; apply ratio to real contexts and problems (such as those involving probability)
- relate ratios to fractions and vice versa
- define percentage as ‘number of parts per hundred’; interpret percentages and percentage changes as a fraction or a decimal, and interpret these multiplicatively; express one quantity as a percentage of another; compare two quantities using percentages

Appendix 3 - statistical enquiry cycle

Specifications must encourage the application of techniques within the framework of the statistical enquiry cycle using real data taken from authentic contexts.

Specifications must require students to understand the importance of planning their line of enquiry or investigation:

- defining a question or hypothesis (or hypotheses) to investigate
- deciding what data to collect and how to collect and record it giving reasons
- developing a strategy for how to process and represent the data giving reasons

Specifications must enable students to recognise the constraints involved in sourcing appropriate data:

- designing collection methods for primary data
- researching sources of secondary data, including from reference publications, the internet and the media
- appreciating the importance of acknowledging sources
- recognising where issues of sensitivity may influence data availability

Specifications must require students to understand the ways that data can be processed, presented and discussed:

- organising and processing their collected data, including via the use of appropriate technologies
- generating diagrams and visualisations to represent the data, using technology and/or manual constructions where appropriate
- generating statistical measures to compare data, using technology and/or manual calculations where appropriate

Specifications must require students to understand that results must be interpreted with reference to the context of the problem:

- analysing/interpreting the diagrams and calculations/measures
- reaching conclusions that relate to the questions and hypotheses addressed
- making inferences and/or predictions
- discussing the reliability of findings

Specifications must require students to present and interpret statistical information that shows and understanding of:

- the clear and concise communication of findings and key ideas
- the appropriate use of technology
- the awareness of the context and target audience

Specifications must enable students to understand the importance of evaluating their own statistical work and that of others:

- identifying weaknesses in approach or representation
- suggesting improvements to processes or the presentation
- refining the process to elicit further clarification of the initial hypothesis

Department for Education

About this publication:

enquiries www.education.gov.uk/contactus

download www.gov.uk/government/publications

Follow us on Twitter: @educationgovuk

Like us on Facebook: facebook.com/educationgovuk