Department for Education

Statistics

Draft GCE AS and A level content

September 2015

Contents

The content for AS and A level statistics

Introduction

1. AS and A level subject content sets out the knowledge, understanding and skills common to all AS and A level specifications in statistics and must require students to apply the statistical techniques listed in paragraph 8 and 9 through the practical investigation of real world data collected from authentic contexts within the framework of the statistical enquiry cycle, as detailed in Appendix 4.

Aims and objectives

2. A level statistics builds upon the statistics and probability components of GCSE mathematics and employs the statistical enquiry cycle to help make sense of data trends and to solve statistical problems in a variety of contexts, such as psychology, biology, geography, business and the social sciences. It prepares students for further study and employment in a wide range of disciplines which use statistical analysis and reasoning with data. The AS statistics is a subset of the A level, which is aimed specifically at facilitating the development of the statistical elements employed across the A level curriculum and supports transition to higher education or employment in any of the many disciplines that make use of quantitative analysis.

3. AS and A level specifications in statistics must encourage students to:

- apply appropriate mathematical and statistical formulae, as set out in appendix 1 and appendix 2
- engage in a practical manner with the statistical enquiry cycle and the research methodologies used in experiments and surveys
- apply statistical techniques using data sourced from a variety of contexts, appreciating when samples or population data could be used and applying appropriate sampling techniques
- generate and interpret the diagrams, graphs and measurement techniques used in performing statistical investigations
- use visualisation of authentic multivariate data to get a qualitative understanding of the multiple factors that interact in real life situations, including, but not limited to, population characteristics, environmental considerations, production variables etc.
- gain direct experience of using some of the technologies that have enabled the collection, visualisation and analysis of large data sets to inform decision-making processes in public, commercial and academic sectors
- develop skills in interpretation and critical evaluation of methodology including justifying the techniques used for statistical problem solving

Detailed content statements

Overview

4. AS and A level statistics specifications assumes a basic level of numerical skills as listed in Appendix 3, as studied in the GCSE (9-1) mathematics subject content. The knowledge and skills required for AS statistics are shown in paragraph 8 with the knowledge and skills required for A level statistics in both paragraphs 8 and 9.

5. AS and A level specifications in statistics must require students to demonstrate an understanding of the statistical enquiry cycle, as detailed in Appendix 4, 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.

6. Specifications must require students to explore a variety of data sets sourced from across the curriculum, from general published sources and from primary data collection activities throughout their course of study. Students must develop the skills required to interpret real data presented in summary or graphical form and use data to investigate questions arising in authentic contexts.

7. Specifications must require students to show knowledge of and become familiar with the ways in which technology can help explore data, in particular mathematical and statistical graphing tools and spreadsheets. Students must be able to calculate summary statistics and access probabilities from standard statistical distributions using manual techniques and to interpret the results obtained through the use of automated functions on handheld or desktop machines.

AS statistics

8. Content required for AS statistics is shown below. This, assessed in the context of the statistical enquiry cycle, represents 100% of the AS content.

A1.1 - numerical measures, graphs and diagrams

- interpret statistical diagrams including bar charts, stem-and-leaf diagrams, box-and-whisker plots, cumulative frequency diagrams, histograms (with either equal or unequal class intervals), time series and scatter diagrams
- know the features needed to ensure an appropriate representation of data using the above diagrams, and how misrepresentation may occur
- justify appropriate graphical representation and comment on those published

- compare different data sets, or the disaggregation of a population by different characteristics, using appropriate diagrams or calculated measures of central tendency and spread: mean, median, mode, range, interquartile range, percentiles, variance and standard deviation
- calculate measures using IT and manually as appropriate
- identify outliers by inspection and using appropriate calculations
- determine the nature of outliers in reference to the population and original data collection process

A1.2 - probability

- know and use language and symbols associated with set theory in the context of probability
- represent and interpret probabilities using tree diagrams, Venn diagrams and two-way tables
- calculate and compare probabilities: single, independent and conditional probabilities
- use permutations and combinations to determine outcomes and probabilities in context
- determine if two events are statistically independent

A1.3 - population and samples

- know and generate both simple (without replacement) and unrestricted (with replacement) random samples
- generate random numbers from various sources and know how to obtain a random sample
- know, demonstrate and evaluate the practical application of random and non-random sampling techniques: simple random, systematic, cluster, judgemental and snowball, including the use of stratification (in proportional and disproportional ratios) prior to sampling taking place
- know the advantages and limitations of sampling methods
- make reasoned choices with reference to the context in which the sampling is to take place, examples include, but not limited to: market research, exit polls, experiments, quality assurance

A1.4 - introduction to probability distributions

- know and use terms for variability: random, discrete, continuous, dependent, independent
- calculate probabilities and determine expected values, variances and standard deviations
- use discrete random variables to model real-world situations
- know the properties of a continuous distribution

- interpret graphical representations of characteristic discrete random variables and continuous distributions such as triangular and rectangular distributions

A1.5 - binomial distribution

- know when a binomial model is appropriate (in real world situations including modelling assumptions)
- know methods to evaluate or read probabilities using formula and tables
- calculate and Interpret the mean and variance

A1.6 - Normal distribution

- know the specific properties of the Normal distribution, and know that data from such an underlying population would approximate to having these properties, with different samples showing variation
- apply knowledge that approximately $\frac{2}{3}$ of observations lie within $\mu \pm \sigma$, and equivalent results for 2σ, and 3σ
- determine probabilities and unknown parameters with a Normal distribution
- apply the Normal distribution to model real world situations
- use the fact that the distribution of $\overline{X}$ has a Normal distribution if X has a Normal distribution

A1.7 - exponential and Poisson distributions

- determine when a Poisson model is appropriate (in real world situations including modelling assumptions)
- determine when an exponential distribution is appropriate (and its relationship to the Poisson distribution)
- evaluate probabilities for Poisson and exponential distributions and know the corresponding mean and variance

A1.8 - introduction to hypothesis testing

- know and use the language of statistical hypothesis testing: null hypothesis; alternative hypothesis; significance level; test statistic; 1-tail test; 2-tail test; critical value; critical region; acceptance region and p-value
- know that a sample is being used to make an inference about the population
- conduct a statistical hypothesis test for the proportion in the binomial distribution and interpret the results in context
- conduct a statistical hypothesis test for the mean of a Normal distribution with known or assumed, from a large sample, variance and interpret the results in context
- know the importance of appropriate sampling when using hypothesis tests and be able to critique the conclusions drawn from rejecting or failing to reject a null hypothesis by considering the test performed

A1.9 - contingency tables

- construct contingency tables from real data, combining data where appropriate and interpret results in context
- use a χ^2 test with the appropriate number of degrees of freedom to test for independence in a contingency table and interpret the results of such a test

A1.10 - unpaired and paired tests

- use sign, Wilcoxon signed-rank tests to investigate population median in single sample tests and also to investigate for differences using a paired model
- use the Wilcoxon rank-sum test to investigate for difference between independent samples

A1.11 - introduction to experimental design

- apply appropriate techniques within the statistical enquiry cycle to investigate authentic situations using real data
- understand the practical constraints of collecting unbiased data
- appreciate that data can be misrepresented when used out of context or through misleading visualisation

A level statistics

9. Content required for A level statistics is shown in the AS statistics paragraphs A1.1-A1.11, plus paragraphs A2.1-A2.11. The content set out in these paragraphs, assessed in the context of the statistical enquiry cycle; represent 100% of the A-level content.

A2.1 - Bayes' theorem

- know conditional probability to include Bayes' theorem for up to three events, including the use of tree diagrams

A2.2 - probability distributions

- know the use and validity of distributions which could be appropriate in a particular real-world situation: binomial, Normal, Poisson and exponential
- evaluate the mean and variance of linear combinations of independent random variables through knowledge that if X_i are independently distributed (μ_i, σ_i^2) then $\sum a_i X_i$ is distributed $(\sum a_i \mu_i, \sum a_i^2 \sigma_i^2)$
- evaluate linear combinations of two or more independent Normal distributions

A2.3 - further experimental design

- know and discuss issues involved in experimental design: experimental error, randomisation, replication, control and experimental groups, blind and double blind trials
- know the use of paired comparisons and blocking to reduce experimental error
- use completely random and randomised block designs

A2.4 - sampling, estimates and resampling

- use and demonstrate understanding of terms parameter, statistic, unbiased and standard error
- know the use of the central limit theorem in the distribution of $\overline{X}$ where the initial distribution, X, is not Normally distributed and the sample is large
- know bootstrapping as a technique used to make statistical inference about a population by resampling a collected sample multiple times. Be able to interpret diagrams representing bootstrap simulations using the central 95% of the re-sample estimates to find the upper and lower limits of an uncertainty interval or interval estimate (bootstrap confidence interval)

A2.5 - hypothesis testing, significance testing, confidence intervals, and power

- interpret and calculate type I and type II errors, in hypothesis testing and know their practical meaning
- use confidence intervals for the mean using *z* or *t* as appropriate, including the difference between two means, interpreting results in practical contexts
- know that a change in sample size will affect the width of a confidence interval
- evaluate the strength of conclusions and misreporting of findings from hypothesis tests, including the calculation and importance of the power of a hypothesis test
- know that sample size can be changed to potentially elicit appropriate evidence in a hypothesis test
- know the difference and advantages of using critical regions or p-values as appropriate in real-life contexts in all tests in these subject content

A2.6 - hypothesis testing for 1 and 2 samples

- know how to apply knowledge about carrying out hypothesis testing to conduct tests for the:
    - mean of a Normal distribution with unknown variance using the *t* distribution
    - difference of two means for two independent Normal distributions with known variances

- difference of two means for two independent Normal distributions with unknown but equal variances
- difference between two binomial proportions

- interpret results for these tests in context

A2.7 - paired tests

- use sign, Wilcoxon signed-rank or paired *t* test, understanding appropriate test selection and interpreting the results in context

A2.8 - goodness of fit

- conduct a statistical goodness of fit test using $\sum \frac{(O-E)^2}{E}$ as an approximate χ^2 statistic

A2.9 - analysis of variance

- know one-way analysis of variance, using a completely randomised design with appreciation of the underlying model with additive effects and experimental errors distributed as N(0, σ^2)
- know two-way analysis of variance without replicates, using a randomised block design with blocking
- identify assumptions and interpretations in context

A2.10 - effect size

- know the notion of effect size as a complementary methodology to standard significance testing, and apply in authentic contexts
- know and use Cohen's d in simple situations
- know that effect size is more useful when only interested in the status of the sample, not the population

A2.11 - correlation and linear regression

- calculate (only using appropriate technology - calculator), and interpret association using Spearman's rank correlation coefficient or Pearson's product moment correlation coefficient
- know the appropriate conditions for the use of each of these methods of calculating correlation and determine appropriate approach to assessing correlation in context
- calculate (only using appropriate technology - calculator) and interpret the coefficients for a least squares regression line in context; interpolation and extrapolation, use of residuals to evaluate the model and identify outliers

Appendix 1

Statistical formulae for A level statistics that students are expected to learn and must not be given in the assessment.

Specification reference	Formula required
A1.1	Calculating the angle for a sector in a pie chart: $\frac{x}{total} \times 360$
A1.1	Frequency density for a histogram: *frequency density* = $\frac{frequency}{classwidth}$
A1.1	Calculation of arithmetic mean: $\bar{x} = \frac{\sum fx}{\sum f}$
A1.1	Range = highest value – lowest value
A1.1	Interquartile range (IQR) = upper quartile – lower quartile
A1.1	Interpercentile range and interdecile range as appropriate: percentile 1 – percentile 2 (calculation of a percentage needed)
A1.1	Identification of an outlier: Small outlier is < LQ – 1.5IQR Large outlier is > UQ + 1.5IQR Outlier is also outside $\mu \pm 3\sigma$
A1.2	Formulae for independent events: P(A∩B) = P(A)P(B) P(A I B) = P(A) and P(B I A) = P(B)
A1.2	$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$ $\mathrm{P}(A \cap B) = \mathrm{P}(A) \times \mathrm{P}(B \mid A)$
A1.2	Formulae for permutations and combinations: $\frac{n!}{(n-r)!}$ and $\frac{n!}{(n-r)!r!}$ respectively
A1.3	In order to carry out stratification, calculating the percentage or proportion of an amount: $\frac{x}{100} \times$ amount
A1.4	Properties of random variables: expectation (mean): $\mathrm{E}(X) = \mu = \sum x_i p_i$ variance: $\mathrm{Var}(X) = \sigma^2 = \sum (x_i - \mu)^2 p_i = \sum x_i^2 p_i - \mu^2 = \mathrm{E}(X^2) - \mu^2$

A2.5	Bootstrapping confidence interval: estimate ± *t* or *z* value × standard error
A2.6	Other confidence intervals: Several variations on estimate ± *t* or *z* value × standard error according to the particular confidence interval

Appendix 2

Statistical formulae for A level statistics that students are expected to be familiar with and will be given in the assessment, either with the question or as a separate formulae sheet or booklet.

Specification Reference	Formula required
A1.1	Population variance = $\frac{1}{N}\sum(x-\mu)^2$ Population standard deviation = $\sqrt{\frac{1}{N}\sum(x-\mu)^2}$
A1.1	Sample variance = $\frac{1}{N-1}\sum(x-\overline{x})^2$ Sample standard deviation = $\sqrt{\frac{1}{N-1}\sum(x-\overline{x})^2}$
A1.5	Binomial probability calculations: $P(X=x)\ \binom{n}{x}p^x(1-p)^{n-x}$ Mean = np Variance = $np(1-p)$
A1.6	Normal probability distribution: $\frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
A1.7	Substitute into appropriate test statistics, know the statistic but be given the distribution: $\mathrm{E}(Y)=p$ and $\mathrm{Var}(Y)=\frac{p(1-p)}{n}$ (test on a proportion) Test statistic for the test on a mean $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\sim \mathrm{N}(0,1)$
A1.8	Formula for chi-squared test $\sum\frac{(O_i-E_i)^2}{E_i}$
A2.1	Bayes' theorem for up to three events: $\mathrm{P}(A_j\|B)=\frac{\mathrm{P}(A_j)\times\mathrm{P}(B\|A_j)}{\sum_{i=1}^{n}\mathrm{P}(A_i)\times\mathrm{P}(B\|A_i)}$

A2.2	The Poisson probability formula $P(X = x) = \mathrm{e}^{-\lambda}\,\frac{\lambda^x}{x!}$ and knowledge of the mean and variance being λ
A2.2	The exponential probability formula: $P(X = x) = \lambda\,\mathrm{e}^{-\lambda x}$ and knowledge of the mean and variance being $\frac{1}{\lambda}$ and $\frac{1}{\lambda^2}$ respectively
A2.3	$\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$ $\mathrm{Var}(aX \pm bY) = a^2\ \mathrm{Var}(X) + b^2\ \mathrm{Var}(Y)$ but E(*X*+*Y*) = E(*X*) + E(*Y*) not given
A2.5	Central limit theorem (distribution of $\overline{X}$): E($\bar{X}$) = μ $\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$
A2.7	Test statistic for a mean using t distribution: $\frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$
A2.7	Test statistic for difference of two independent Normal means with known variances: $\frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \sim \mathrm{N}(0,1)$
A2.7	Test statistic for difference of two independent Normal means with unknown but equal variance: $\frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{\sqrt{S_p^2\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}} \sim t_{\,n_x + n_y - 2}$ where $S_p^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}$
A2.7	Test statistic for the difference in two binomial proportions: $\frac{p_1 - p_2}{s.error}$ where s. error = $\sqrt{p \times (1-p) \times (\frac{1}{n_1} + \frac{1}{n_2})}$ where $p = \frac{p_1 \times n_1 + p_2 \times n_2}{n_1 + n_2}$

A2.8	Paired t-test: $\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$
A2.9	Goodness of fit test: $\sum \frac{(O_i - E_i)^2}{E_i}$
A2.10	Analysis of variance (one-way and two-way): one-factor model $x_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, where $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$ total sum of squares $SS_T = \sum_i \sum_j x_{ij}^2 - \frac{T^2}{n}$ between groups sum of squares $SS_B = \sum_i \frac{T_i^2}{n_i} - \frac{T^2}{n}$ two-factor model (with m rows and n columns) $x_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$, where $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$ total sum of squares $SS_T = \sum_i \sum_j x_{ij}^2 - \frac{T^2}{mn}$ between rows sum of squares $SS_R = \sum_i \frac{R_i^2}{n} - \frac{T^2}{mn}$ between columns sum of squares $SS_C = \sum_j \frac{C_j^2}{m} - \frac{T^2}{mn}$
A2.11	Cohen's d formula: $d = (\bar{x}_1 - \bar{x}_2)/s$ where $s = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$
A2.12	Spearman's correlation coefficient: $r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$
A2.12	Product moment correlation coefficient: $r = \frac{S_{xy}}{\sqrt{S_{xx} \times S_{yy}}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\{\sum (x_i - \bar{x})^2\}\{\sum (y_i - \bar{y})^2\}}} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sqrt{\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)\left(\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right)}}$
A2.12	Coefficients for least squares regression line: least squares regression line of y on x is $y = a + bx$, where $a = \bar{y} - b\bar{x}$

	the regression coefficient of y on x is $b = \frac{S_{xy}}{S_{xx}} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$

Appendix 3

The numerical skills, developed in GCSE (9-1) mathematics, required for A level 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)
- recognise, use and manipulate numbers in 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
- solve algebraic equations, including simultaneous equations
- 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 including working out a unit for a rate
- 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 4

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:

- identifying factors that may be related to the problem under investigation
- defining a question or hypothesis (or hypotheses) to investigate
- deciding what data to collect and how to collect and record it giving reasons
- engaging in exploratory data analysis in order to investigate the situation
- developing a strategy for how to process and represent the data giving reasons
- justifying the proposed plan with regards ensuring a lack of bias

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

- designing unbiased collection methods for primary sample data
- obtaining appropriate data from secondary sources including from the internet
- recognise the importance of declaring the data collection methodology, including the need to acknowledge sources
- appreciating the inherent bias that may be incorporated through the use of leading questions either by accident or through agenda driven design

Students must understand a range of techniques in order to process, represent and discuss data:

- organising and processing the collected data including via the use of appropriate technologies
- make inferences about the population using appropriately chosen diagrams and summary measures to represent the data including via the use of appropriate technologies
- appreciating how to avoid misrepresentation of the data

Students must appreciate the need to consider the context of the problem when interpreting results:

- analysing/interpreting diagrams and calculations/measures
- drawing together conclusions that relate to the questions and hypotheses addressed
- using appropriate tests to determine the statistical significance of the findings
- discussing the reliability of their and others' findings

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

- the clear and concise communication of findings and key ideas

- the appropriate use of technology
- an awareness of the context and target audience

Students must be able to evaluate their own statistical work and that of others:

- identifying weaknesses in approach used to collect or display data
- recognising the limitations of the findings by considering the sample size and sampling technique
- suggesting improvements to the statistical processes used
- refining processes 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