



Higher National Unit Specification

General information

Unit title: Working with Data (SCQF level 7)

Unit code: J4Y4 34

Superclass: CA

Publication date: December 2020

Source: Scottish Qualifications Authority

Version: 02

Unit purpose

The purpose of this unit is to introduce learners to data analysis. It will provide essential analysis skills to data workers, who want to acquire data skills as part of their current work role. The focus of the unit is data preparation and analysis, rather than data visualisation and communication.

This **non-specialist** unit is designed for learners with no previous experience of data analysis, although numeracy skills and familiarity with computer software are assumed. Previous experience of spreadsheets is desirable but not essential.

Learners will understand the data analysis process and how to use data analysis tools to extract, transform and analyse data. Basic statistical methods (such as correlation) will be introduced in context. Learner will gain insights into large datasets using a range of data analysis tools and techniques to facilitate data-driven decision making.

At the completion of the unit, learners will be able to carry out analyses of familiar data to aid decision-making. Learners may progress to J4Y5 34 *Communicating with Data* at SCQF level 7 or J4Y6 35 *Working with Data* at SCQF level 8.

Outcomes

On successful completion of the unit the learner will be able to:

- 1 Extract data from different sources.
- 2 Transform data to prepare for analysis.
- 3 Analyse data to provide insights.

Higher National Unit Specification: General information (cont)

Unit title: Working with Data (SCQF level 7)

Credit points and level

1 Higher National Unit credit at Scottish Credit and Qualifications Framework (SCQF) level 7:
(8 SCQF credit points at SCQF level 7)

Recommended entry to the unit

This non-specialist unit is designed for learners with no previous experience of data analysis. However, numeracy skills are assumed; it is recommended learners possess the Core Skill in *Numeracy* at SCQF level 5 or above. A familiarity with spreadsheet software is desirable.

Core Skills

Achievement of this Unit gives automatic certification of the following Core Skills component:

Core Skill component	Critical Thinking at SCQF level 6
Complete Core Skill	Information and Communication Technology at SCQF level 6

There are also opportunities to develop aspects of Core Skills which are highlighted in the support notes of this unit specification.

Context for delivery

If this unit is delivered as part of a group award, it is recommended that it should be taught and assessed within the subject area of the group award to which it contributes.

This unit is complimentary to J4Y5 34 *Communicating with Data* at SCQF level 7, which focuses on the visualisation of data. As such, it is recommended that teaching, learning and assessment are combined with that unit when possible.

Equality and inclusion

This unit specification has been designed to ensure that there are no unnecessary barriers to learning or assessment. The individual needs of learners should be taken into account when planning learning experiences, selecting assessment methods or considering alternative evidence.

Further advice can be found on our website www.sqa.org.uk/assessmentarrangements.

Higher National Unit Specification: Statement of standards

Unit title: Working with Data (SCQF level 7)

Acceptable performance in this unit will be the satisfactory achievement of the standards set out in this part of the unit specification. All sections of the statement of standards are mandatory and cannot be altered without reference to SQA.

Where evidence for outcomes is assessed on a sample basis, the whole of the content listed in the knowledge and/or skills section must be taught and available for assessment. Learners should not know in advance the items on which they will be assessed, and different items should be sampled on each assessment occasion.

Outcome 1

Extract data from different sources.

Knowledge and/or skills

- ◆ Internal and external data sources
- ◆ Types of data (categorical and numerical data and their sub-types)
- ◆ Common data formats
- ◆ Data quality including data bias
- ◆ Tools for extracting and transforming data

Outcome 2

Transform data to prepare for analysis.

Knowledge and/or skills

- ◆ Types of data transformation
- ◆ Common transformations including filtering, sorting, combining, separating and grouping
- ◆ Data cleaning
- ◆ Data loading
- ◆ Legal and ethical considerations

Outcome 3

Analyse data to provide insights.

Knowledge and/or skills

- ◆ Types of software for data analysis
- ◆ Descriptive and diagnostic analytics
- ◆ Data structures including lists and tables
- ◆ Data aggregation including pivot tables
- ◆ Normal distribution
- ◆ Summary statistics
- ◆ Measures of central tendency
- ◆ Measures of dispersion including standard deviation
- ◆ Concept of correlation and correlation co-efficient
- ◆ Significance of domain knowledge in data analysis

Higher National Unit Specification: Statement of standards (cont)

Unit title: Working with Data (SCQF level 7)

Evidence requirements for this unit

The evidence requirements for this unit will take **one** form: **product** evidence.

The product evidence will consist of **at least one** analysed dataset. The analysis must relate to real data. The analysis may be descriptive or diagnostic.

The dataset should be extracted from at least two sources. The dataset may be familiar to learners but must comprise at least 1,000 multi-variate records, which require significant cleaning. The dataset must include a range of data types, including dates and text, some of which will require categorisation prior to analysis.

Learners must carry out a range of transformations on the data including resolving data types, combining and splitting data, and removal of unnecessary data. The transformed dataset must be properly structured, suitable for analysis, and efficient in terms of data storage. The dataset must be legal and ethical. Evidence must be provided that data bias has been considered.

The analysis of the dataset must include at least one pivot table to summarise the data, various measures of spread and dispersion, including standard deviation, and must identify all significant summaries and relationships in the data, including correlations where these exist. The results of the analysis must provide insights into the dataset.

The evidence may be produced over an extended period of time, under loosely controlled conditions.

The SCQF level of this unit provides additional context on the nature of the required evidence and the associated standards. The level descriptors should be used (explicitly or implicitly) when making judgements about the evidence.

When evidence is produced in loosely controlled conditions it must be authenticated. The *Guide to assessment* provides further advice on methods of authentication.

The *Guidelines on approaches to assessment* (see the support notes section of this specification) provides specific examples of instruments of assessment.



Higher National Unit Support Notes

Unit title: Working with Data (SCQF level 7)

Unit support notes are offered as guidance and are not mandatory.

While the exact time allocated to this unit is at the discretion of the centre, the notional design length is 40 hours.

Guidance on the content and context for this unit

This unit is designed to introduce learners to the key skills, and underpinning knowledge, needed to analyse data. Although there are some theoretical aspects, the focus of the unit should be learning to work with data in a practical way.

A variety of software tools could be used in this unit. It is recommended that, when possible, learners are exposed to a range of software. For example, the unit could be delivered (entirely) using Microsoft Excel™ but the data transformations would be best done using alternative software (such as Query Editor™), and a combination of products is recommended.

The scope of this unit does not include data visualisation. The outcomes relate to the ETL (Extract, Transform, Load) process only. For a more complete learning experience, learners should undertake this unit alongside J4Y5 34 *Communicating with Data* at SCQF level 7.

Whenever possible, learning should take place using real data. This data should be familiar to the learner and, ideally, relevant to the learner's current work role. Datasets should vary in their size and 'messiness' but, at some stage in the unit, learners should experience large, messy datasets.

Please note that the following guidance, relating to specific outcomes, does not seek to explain each knowledge/skills statement, which is left to the professionalism of the teacher. It seeks to clarify the *Statement of Standards* where it is potentially ambiguous. It also focuses on non-apparent teaching and learning issues that may be over-looked, or not emphasised, during unit delivery. As such, it is not representative of the relative importance of each knowledge/skill.

Outcome 1

This outcome relates to data extraction. The knowledge and skills statements are self-explanatory. Data quality is an important part of this outcome. It is vital that learners appreciate the critical importance of data quality in data analysis. The discussion about data bias ('bias in/bias out') should emphasise the problem of historical data bias and include ways of addressing this problem.

Higher National Unit Support Notes (cont)

Unit title: Working with Data (SCQF level 7)

Outcome 2

This outcome relates to data transformation. Emphasis should be placed on the importance, (and time-consuming nature) of this stage in the data analysis process. At this level, the required transformations should be non-complex, but learners should gain experience of cleaning large, messy datasets.

This outcome also covers the legal and ethical aspects of data analysis. With regards to ethics, learners should be made aware of the difficulty of protecting the privacy of data subjects due to the inherent problem of ensuring anonymity in datasets. Methods of improving data privacy should be explored (such as differential privacy). There is scope in this outcome to discuss the tension between fairness and accuracy — how increasing privacy (through anonymising the data or not recording certain attributes such as race or gender) can reduce the descriptive or predictive accuracy of the analysis.

Outcome 3

This outcome relates to data analysis. Statistical concepts and methods should be basic. The most complex techniques are standard deviation and correlation. Correlation may be limited to the Pearson correlation co-efficient(s). It is important that learners understand the underlying statistical concepts and not merely compute statistical values. The distinction between descriptive analytics and diagnostic analytics should be understood.

Guidance on approaches to delivery of this unit

A suggested distribution of time, across the outcomes, is:

Outcome 1: 10 hours

Outcome 2: 15 hours

Outcome 3: 15 hours

Learners will require access to appropriate hardware and software throughout this unit. A range of software could be used to facilitate learning (see *Guidance on the content and context of this unit*).

This is a practical unit and the focus should be on the acquisition of practical skills in data analysis. However, theory should be introduced in context. For example, in outcome 1, the range of data formats should include a variety of common data formats, including those not likely to be used in the learner's work role. In the early stages of this unit, there is an opportunity to expose learners to the typical uses of data analysis using contemporary case studies.

When learners are in employment, they should be given the opportunity to work with datasets relevant to their work roles. Irrespective of their employment status, use should be made of familiar data that learners can relate to. The datasets used for learning should vary in their size and complexity. At some point in the unit, learners should be exposed to realistic datasets in terms of size and messiness.

Higher National Unit Support Notes (cont)

Unit title: Working with Data (SCQF level 7)

Guidance on approaches to assessment of this unit

Evidence can be generated using different types of assessment. The following are suggestions only. There may be other methods that would be more suitable to learners.

The evidence requirements for this unit may be satisfied in two ways. An assignment may be set, towards the end of the unit, requiring learners to carry out a specific analysis on a familiar dataset. This single, practical activity would have to satisfy all the evidence requirements. Alternatively, learners could maintain a portfolio of analyses carried out during the life of the unit. The portfolio would collectively satisfy all the evidence requirements (but note that at least one analysis must comprise at least 1,000 records).

There are opportunities to carry out formative assessment at various stages in the unit. For example, formative assessment could be carried out on the completion of each outcome to ensure that learners have grasped the knowledge contained within it. This would provide assessors with an opportunity to diagnose misconceptions, and intervene to remedy them before progressing to the next outcome.

Centres are reminded that prior verification of centre-devised assessments would help to ensure that the national standard is being met. Where learners experience a range of assessment methods, this helps them to develop different skills that should be transferable to work or further and higher education.

Opportunities for e-assessment

E-assessment may be appropriate for some assessments in this unit. By e-assessment we mean assessment which is supported by Information and Communication Technology (ICT), such as e-testing or the use of e-portfolios or social software. Centres which wish to use e-assessment must ensure that the national standard is applied to all learner evidence and that conditions of assessment as specified in the evidence requirements are met, regardless of the mode of gathering evidence. The most up-to-date guidance on the use of e-assessment to support SQA's qualifications is available at www.sqa.org.uk/e-assessment.

Opportunities for developing Core and other essential skills

Opportunities to develop aspects of Core Skills occur in all outcomes. These are:

Information and Communication Technology (ICT) at SCQF level 6

Numeracy at SCQF level 6

Learners will learn, and apply, *ICT* skills throughout this unit through their use of computer hardware and software required to carry out analyses.

Learners will also develop their numeracy knowledge and skills in outcome 3 when they use a range of arithmetical and statistical functions to perform analyses.

Higher National Unit Support Notes (cont)

Unit title: Working with Data (SCQF level 7)

Guidance on approaches to assessment of this unit

The Critical Thinking component of Problem Solving at SCQF level 6 is embedded in this unit. When a learner achieves the unit, their Core Skills profile will also be updated to include this component.

The Core Skill of Information and Communication Technology at SCQF level 6 is embedded in this unit. When a learner achieves the unit, their Core Skills profile will also be updated to include this Core Skill.

History of changes to unit

Version	Description of change	Date
02	Core Skills Component Critical Thinking at SCQF level 6 embedded. Core Skill Information and Communication Technology at SCQF level 6 embedded.	09/12/20

© Scottish Qualifications Authority 2020

This publication may be reproduced in whole or in part for educational purposes provided that no profit is derived from reproduction and that, if reproduced in part, the source is acknowledged.

Additional copies of this unit specification can be purchased from the Scottish Qualifications Authority. Please contact the Business Development and Customer Support team, telephone 0303 333 0330.

Unit template: June 2017

General information for learners

Unit title: Working with Data (SCQF level 7)

This section will help you decide whether this is the unit for you by explaining what the unit is about, what you should know or be able to do before you start, what you will need to do during the unit and opportunities for further learning and employment.

This unit will introduce you to data analysis. No previous knowledge of data analysis is assumed. You will learn from the beginning. A familiarity with spreadsheet software is desirable but not essential.

You will use a variety of software to carry out your analyses, such as Microsoft Excel™ and Query Editor™. You will use this software to practise with real data to answer real business questions. The focus of this unit is the acquisition of practical skills in data analysis so that you can make data-driven decisions in your current work role.

The unit covers a variety of topics including:

- ◆ ETL (Extract/Transform/Load) process
- ◆ types of data
- ◆ data formats
- ◆ data quality
- ◆ data transformations
- ◆ data cleaning
- ◆ data structures
- ◆ basic statistics
- ◆ descriptive and diagnostic analytics
- ◆ using software to perform these analyses
- ◆ data ethics

During the unit, you will work with datasets relevant to your current work role. You will learn how to use data analysis to gain insights into these datasets and make data-driven decisions based on your analysis.

The unit may be assessed by a practical activity (involving you in carrying out a complete analysis) or by maintaining a portfolio of your work throughout the unit.

This unit may be taken alongside J4Y5 34 *Communicating with Data* at SCQF level 7, which focuses on the visualisation and communication of the results of your analyses.

The Critical Thinking component of Problem Solving at SCQF level 6 is embedded in this unit. When a learner achieves the unit, their Core Skills profile will also be updated to include this component

The Core Skill of Information and Communication Technology SCQF at level 6 is embedded in this unit. When a learner achieves the unit, their Core Skills profile will also be updated to include this Core Skill.