

2021

A case study in the e-assessment of statistics for non-specialists

Iain Weir

University of the West of England, England, iain.weir@uwe.ac.uk

Rhys Gwynllyw

University of the West of England, England, rhys.gwynllyw@uwe.ac.uk

Karen Henderson

University of the West of England, England, karen.henderson@uwe.ac.uk

Follow this and additional works at: <https://ro.uow.edu.au/jutlp>

Recommended Citation

Weir, I., Gwynllyw, R., & Henderson, K. (2021). A case study in the e-assessment of statistics for non-specialists. *Journal of University Teaching & Learning Practice*, 18(2). <https://doi.org/10.53761/1.18.2.5>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

A case study in the e-assessment of statistics for non-specialists

Abstract

The need to be able to choose, perform and interpret the results from appropriate statistical tests is ubiquitous in many STEM disciplines and beyond. In this case study, we illustrate a learning and assessment approach that has been used successfully for first year Business School students taking a module on Business Decision Making for Marketing and Events. The objective of the module is to give students a grounding in statistics, including learning to use the SPSS package, and to apply the techniques learnt to a large real-world data set. Due to the large number of students and their diverse abilities the learning and teaching strategy is centred on using formative key skill e-assessments to encourage students to self-learn. On accessing each e-assessment, students are presented with a randomly generated data set, which they are required to import into SPSS to then appropriately analyse and report on. Marking and feedback occurs automatically on submission and repeated use ensures that a student thoroughly learns a key skill and covers various analysis outcome scenarios, for instance significant or not significant test outcomes. Each key skill e-assessment has multiple embedded links to comprehensive Help pages that provide SPSS 'how-to' information or output interpretation. Access to these means that staff in the PC sessions can concentrate on giving higher-level support as opposed to merely helping with the mechanics of producing SPSS output. The e-assessments were created using Dewis, which is a fully algorithmic open-source e-assessment system. Dewis' ability to communicate with the R programming language facilitates the creation of authentic e-assessments, generating bespoke student data and providing answers that match SPSS screen output. The assessment of students' statistical competency is by in-class e-exams, which are summative and sat under exam conditions. The fact that the formative and summative e-assessments are marked immediately by Dewis represents a huge saving in marking time compared to human marking and allows students to have instant feedback on their work. Repeated use of the formative e-assessments allows students to independently practice and check their understanding. In this case study, we find that those students that did practice achieved significantly higher marks in the e-exams. We are now in our third year of delivery and the module has developed further due to stronger connections between the business and statistics strands of the module.

Keywords

Case study, e-Assessment, statistics



2021

A case study in the e-assessment of statistics for non-specialists

Iain Weir

University of the West of England, England, iain.weir@uwe.ac.uk

Rhys Gwynllyw

University of the West of England, England, rhys.gwynllyw@uwe.ac.uk

Karen Henderson

University of the West of England, England, karen.henderson@uwe.ac.uk

Follow this and additional works at: <https://ro.uow.edu.au/jutlp>

Recommended Citation

Weir, I., Gwynllyw, R., & Henderson, K. (2021). A case study in the e-assessment of statistics for non-specialists. *Journal of University Teaching & Learning Practice*, 18(2). <https://ro.uow.edu.au/jutlp/vol18/iss2/05>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

A case study in the e-assessment of statistics for non-specialists

Abstract

The need to be able to choose, perform and interpret the results from appropriate statistical tests is ubiquitous in many STEM disciplines and beyond. In this case study, we illustrate a learning and assessment approach that has been used successfully for first year Business School students taking a module on Business Decision Making for Marketing and Events. The objective of the module is to give students a grounding in statistics, including learning to use the SPSS package, and to apply the techniques learnt to a large real-world data set. Due to the large number of students and their diverse abilities the learning and teaching strategy is centred on using formative key skill e-assessments to encourage students to self-learn. On accessing each e-assessment, students are presented with a randomly generated data set, which they are required to import into SPSS to then appropriately analyse and report on. Marking and feedback occurs automatically on submission and repeated use ensures that a student thoroughly learns a key skill and covers various analysis outcome scenarios, for instance significant or not significant test outcomes. Each key skill e-assessment has multiple embedded links to comprehensive Help pages that provide SPSS 'how-to' information or output interpretation. Access to these means that staff in the PC sessions can concentrate on giving higher-level support as opposed to merely helping with the mechanics of producing SPSS output. The e-assessments were created using Dewis, which is a fully algorithmic open-source e-assessment system. Dewis' ability to communicate with the R programming language facilitates the creation of authentic e-assessments, generating bespoke student data and providing answers that match SPSS screen output. The assessment of students' statistical competency is by in-class e-exams, which are summative and sat under exam conditions. The fact that the formative and summative e-assessments are marked immediately by Dewis represents a huge saving in marking time compared to human marking and allows students to have instant feedback on their work. Repeated use of the formative e-assessments allows students to independently practice and check their understanding. In this case study, we find that those students that did practice achieved significantly higher marks in the e-exams. We are now in our third year of delivery and the module has developed further due to stronger connections between the business and statistics strands of the module.

Keywords

Case study, e-Assessment, statistics

Introduction

Being able to choose and perform appropriate statistical tests and to interpret the results generated from them are key skills required by many disciplines. There is a wealth of published articles on different approaches to and the challenges of teaching statistics as evidenced, for example, by a journal dedicated to statistical education (Journal of Statistical Education, 2020). The challenge of teaching statistics to non-specialists is well-established (Horton, 2015; Mustafa, 1996; Simpson, 1995). Pedagogic approaches that have been used elsewhere include making the statistics relevant and by enabling active learning (STEM learning, 2009; Idris, 2018). It is now relatively straightforward to include multimedia to create blended learning resources and these have proved popular with students (Willis et al., 2018). Wood et al. (2018, p. 10) recommend that students 'should learn to use a statistical software package if possible' as opposed to computing statistical quantities by hand. SPSS (IBM Corp., 2019) is a widely used program for statistical analysis in industry, commerce, banking, local and national government. Consequently, SPSS is often used in university level courses which involve statistical analyses (see for example Aljandali, 2016).

With technological advances, there has been increased demand and use of electronic assessment in education (Brink & Lautenbach, 2011). This not only reduces marking load for academic staff but also is a mechanism for students to receive prompt feedback on their work. The majority of statistics e-assessments use multiple choice questions (MCQs) or involve short calculations that can be done by hand or using Excel (e.g. Blanco & Ginovart, 2010; Delmas et al., 2007; Garfield, 2006; Hodgson & Pang, 2012). Gwynllyw et al. (2015) have developed e-assessments that automatically assess students' ability to identify and perform appropriate statistical analyses in SPSS. The e-assessments are run on Dewis, which is a fully algorithmic open-source e-assessment system. Dewis uses embedded R code in order to create random data sets with desired statistical properties and to generate results.

Using e-assessment for coursework has become standard practice in many institutions (Sangwin, 2013) and it seems likely that online examinations will become standard practice in the near future (Kuikka et al., 2014). However, progress has been slower than predicted by Collins et al. (2003). Gray et al. (2016) found that there is considerable interest in online exams yet few reliable sources of good practice. Roads (2016) reports that most organisations expect to increase their use of e-assessment over the next five years and the Covid-19 pandemic (World Health Organization, 2020) is likely to accelerate its use (Bothwell, 2020).

In this paper we provide a model for how online testing of statistics can be achieved by reporting on our experience of running a first year Business Decision Making (BDM) module taken by over 350 Business School students at the University of the West of England (UWE Bristol) and partner institutions (Villa College, Maldives; British College Kathmandu, Nepal; and Northshore College, Sri Lanka). The module, which ran for the first time in the 2017-18 academic year, comprises a single semester course covering statistical methodology. The material is taught in greater depth than had been achieved previously and has been modernised to include analyses using the statistical software SPSS rather than by hand calculations. In particular, we detail how we have used e-assessments in the module. These take the form of formative key skills e-assessments, which are designed primarily to aid students' learning and in-class summative e-assessments, which we refer to as e-exams, which are designed to efficiently assess students' performance in the module and are run under exam conditions. In preparation for the e-exams students are given access to formative practice e-tests.

Background

Context

The BDM module covers the following statistics topics:

- Exploratory data analysis, including normality testing;
- One sample t-test & nonparametric equivalent;
- Probability & decision trees;
- Two sample t-tests & nonparametric equivalents;
- Critical path analysis & Gantt charts;
- Correlation and simple linear regression;
- Chi-squared tests;
- Time series.

The syllabus presented above constitutes a challenging amount of material to cover, especially since the student cohort is a large, diverse group of non-mathematicians. In addition, students are required to learn how to use SPSS and use the knowledge gained to complete a piece of coursework which is a component of assessment in the module. Further, it is important that students gain sufficient competency/understanding of the techniques so that they are able to perform data analysis when collecting their own primary data for a research project module taken in the following year of their studies. Prior to designing the way the statistics material was to be taught and assessed, several challenges were identified by the module team. These were:

How to teach the students so that they are able to learn SPSS independently;

1. How to get students to attain the required statistical competency and understanding;
2. How to support students to be capable of formal report writing;
3. How to assess the large number of students in a time-efficient way;
4. How to monitor student progress in a large cohort and to target those that are struggling.

Creation of authentic e-assessments

E-assessments were created using Dewis which is a fully algorithmic open-source e-assessment system designed and developed at UWE Bristol. Dewis is a completely stand-alone web-based system used for both summative and formative assessments (Gwynllyw & Henderson, 2009; Dewis Development Team, 2012) initially targeted at assessments in mathematics. The algorithmic nature of each e-assessment enables the system's marking and feedback to respond dynamically to the student's input which facilitates intelligent marking, such as detecting and reporting on common student errors (Gwynllyw & Henderson, 2012; Sikurajapathi et al., 2019). Student access to the system is via the institution's virtual learning environment (VLE) using Learning Tools Interoperability (LTI) protocol for single-sign-on access. Using LTI protocol allows Dewis to be launched from the VLE without students having to log in again. Also, it allows students' marks from e-assessments to be automatically passed back to the VLE.

Dewis' ability to communicate with the R programming language (R Core Team, 2018) greatly facilitates the task of generating bespoke student data and providing answers that match SPSS screen output (Gwynllyw et al., 2015; Weir et al., 2015). A Dewis-R question-authoring interface has been developed to enable the composition of questions using a simplified Dewis question structure with embedded R code that can take advantage of standard R routines (Weir & Gwynllyw, 2017). This simplification is possible since the statistical computations are performed by R and not Dewis. As such, Dewis supplies the administration of the assessment but not the computations. This pioneering approach facilitates the creation of authentic e-assessments, enabling Dewis to generate random bespoke data sets with desired statistical properties that cover a variety of analysis experiences. Furthermore, the statistical results for each data set are generated by R coding, enabling automatic

marking of students' analyses of these data sets and appropriate feedback supplied back to the students immediately upon submission.

Figure 1 shows the activity flow for a student taking a typical Dewis statistics e-assessment. On accessing the e-assessment students download their dataset as a CSV file which they then import into SPSS. Once in SPSS they then perform the necessary operation(s) on their dataset in SPSS to generate results. Finally, they can log back into Dewis to submit their answer(s). On submission, their answers are marked automatically against those generated by the Dewis-R code and, if allowed by the academic, students can view their marks and feedback. Further details of some of the e-assessments used on this module can be found later in the paper.

Teaching and assessment strategy

Overview

Over the twelve-week semester, the statistics staff deliver a one-hour lecture and a one-hour computer practical to students each week. Concurrently, business school staff supply relevant context to each of the statistical topics taught through a one-hour tutorial per week. The business school staff are also responsible for the final summative assessment which is a piece of coursework that students complete in their own time.

At the start of the module, the statistics staff provide a booklet of notes that covers statistics and the use of SPSS. In the lecture, material contained within this booklet is explained. The first computer practical sessions concerns the basics of SPSS (e.g. data entry, variable labels, file opening and saving) and the creation of summary statistics, graphics and tables. For this purpose, several bespoke self-learn "How to" videos were created so that the students could pause at will and thus work independently and at their own pace. This supported active learner engagement and was beneficial due to the wide range of computer abilities amongst the cohort. From the tutors perspective this effortlessly allowed them to accommodate students that had missed the first PC lab session who would otherwise have been a drain on teaching resources.

From the second computer practical session onwards, the main learning tools are Dewis key skills e-assessments. Each time an e-assessment is started Dewis automatically generates a new random data set, which requires the student to import the data into SPSS, analyse appropriately and report on their findings. Repeated use ensures that a student thoroughly learns a key skill and covers various analysis outcome scenarios, for instance significant or not significant test outcomes. The next section of the paper contains further details of one of the key skill e-assessments that is employed on this module.

To supplement the key skills e-assessments, students can download pre-prepared word templates designed to represent a complete statistical analysis on a specific supplied data set. An example of part of one of these pre-prepared word templates is included in Appendix 1. Each template has a similar format; the SPSS output is removed, numerical values are blanked out and multiple choice interpretation decisions to make are inserted. The instructions in the template are clear as to what students are required to do and what they should include in order to complete the template. In most PC lab sessions, students have one or two of these templates to work with. In some cases students are required to download data from the VLE into SPSS and in others the data are small enough that they are able to enter it into SPSS directly (as is the case in the example shown in Appendix 1). Students are instructed to create the required SPSS output and paste it into the relevant point in the template. Three asterisks are used to indicate where numerical values, taken from the SPSS output, have been blanked out and need to be entered. Bold text is used to indicate where students have to

choose the appropriate inference based upon their supporting output. One advantage of this approach is that students have access to a well-structured statistical report format, so can concentrate on statistical data creation and interpretation as opposed to spending time on the structure of the report. Correctly completed templates are available from the VLE once all of the PC lab sessions for that week have occurred. These templates are designed for students to master a methodical and defensible approach to carrying in depth and appropriate statistical analysis. Students are encouraged to download the portfolio of completed templates once they have completed the module. This is in anticipation that, for some students, the portfolio will be useful as a reference resource for any further statistical work they might need to do at university or once they have graduated.

Figure 1

Activity flow for a student taking a Dewis statistics e-assessment.

The Question.
The average IQ in the general population is 100.
The boxplot below is of the IQs of a random sample of people who are taking a certain medication.
Does the 95% confidence interval for the mean indicate that the medication is affecting IQ?

Data and transfer check
Download the [data](#) and transfer it into SPSS. [Help](#)
Check that your data transfer has been successful by obtaining the Std. E should appear in SPSS output as **2.066**. [Help](#)
If you do not have this exact value, then you may have not transferred your data correctly. Do not continue with the test until your value agrees as otherwise.
Unless otherwise directed you should report all numeric values to the accuracy supplied when your data has been transferred correctly.

Question [Help](#)
Complete the following:
The sample of people had a mean IQ of (95% CI: to).
Based upon the 95% confidence interval, it appears that the mean IQ is whilst on the medication.

Marking scheme
• Numerical entries: correct = 1 mark / incorrect = 0 marks
• Dropdown choice: correct = 1 mark / incorrect = 0 marks
• Not answered = 0 marks

Descriptives

	Statistic	Std. Error
IQ Mean	105.40	2.066
95% Confidence Interval for Mean	Lower Bound 101.17 Upper Bound 109.63	
5% Trimmed Mean	105.11	
Median	106.50	
Variance	128.041	
Std. Deviation	11.316	
Minimum	85	
Maximum	131	
Range	46	
Interquartile Range	16	
Skewness	.238	.427
Kurtosis	-.208	.833

Key skill e-assessments

Formative, key skills e-assessments are the main learning tools for students to use in the PC classes. Each key skill e-assessment has multiple embedded links to comprehensive Help pages that provide SPSS ‘how-to’ information or output interpretation. This form of support is more targeted and immediate than a student having to refer back to the notes. It allows students to work independently and at their own pace whilst also freeing up staff to concentrate on giving higher-level interpretative advice.

The e-assessments allow instantaneous and dynamic colour-coded feedback specific to student inputs. For example, feedback is given for inappropriate test choices as well as for incorrect analyses. Green text indicates that the answer was correct and red text indicates that the answer was incorrect. This is an important feature since research has shown that students learn from e-assessment feedback, using it to perfect their technical knowledge (Greenhow & Gill, 2008). We believe the approach of short key skills e-assessments is a very useful learning aid that supports and prepares the student for more in depth e-testing of complete analyses. The wording used in our e-assessments is appropriate for the formal reporting of statistical analyses, which stands students in good stead should they be required to perform statistical analyses in their future careers.

In order to illustrate these features, snapshots from a specific key skill e-assessment are contained in Figures 2-4, which show the presentation, marking and feedback as viewed by a student. The data scenario for this particular key skill e-assessment concerns using a 95% confidence interval (CI) for the mean to infer whether the mean IQ of people on a particular medication appears to differ to the general population IQ value of 100. Figure 2 shows a screenshot of the question that is presented to the student.

The data are supplied to the student via a CSV file as a downloadable link. The successful transfer of the data into SPSS is checked by the student comparing the stated standard error of the mean to that which they obtain in SPSS. Having confirmed agreement, the student is instructed to input all numerical answers to the same number of decimal places that SPSS reports to. However, the e-assessment author could change the desired precision if they wish.

After the appropriate analysis in SPSS, the student is then required to input the numerical values for the sample size, sample mean and the 95% confidence interval limits before choosing from a dropdown menu the correct inference concerning the problem posed. In Figure 2, for illustration purposes, the student’s attempt has a mixture of correct, incorrect and blank answers.

Figure 2

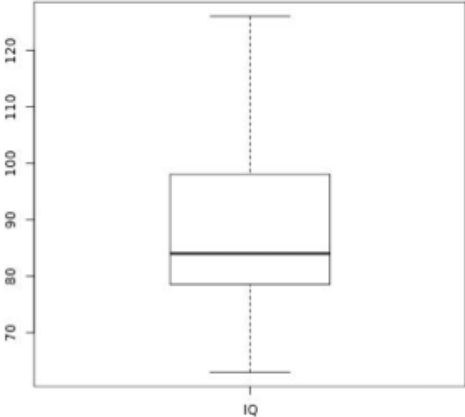
Screenshot of a key skill e-assessment example. Here, the student has (a) entered three correct numerical answers, (b) not supplied the 95% CI upper limit and (c) has an incorrect inference concerning the mean IQ.

The Question.

The average IQ in the general population is 100.

The boxplot below is of the IQs of a random sample of people who are taking a certain medication.

Does the 95% confidence interval for the mean indicate that the medication is affecting IQ?



Data and transfer check

Download the [data](#) and transfer it into SPSS. [Help](#)

Check that your data transfer has been successful by obtaining the Std. Error of the mean for your data which should appear in SPSS output as **4.718**. [Help](#)

If you do not have this exact value, then you may have not transferred your data from the Excel file to SPSS correctly. Do not continue with the test until your value agrees as otherwise you may not have correct answers.

Unless otherwise directed you should report all numeric values to the accuracy displayed in the SPSS output that is supplied when your data has been transferred correctly.

Question [Help](#)

Complete the following:

The sample of people had a mean IQ of (95% CI: to).

Based upon the 95% confidence interval, it appears that the mean IQ is

Select
 Select
 still the same
 higher
 lower

 whilst on the medication.

Marking scheme

- Numerical entries: correct = 1 mark / incorrect = 0 marks
- Dropdown choice: correct = 1 mark / incorrect = 0 marks
- Not answered = 0 marks

Annotations:

- (a) Correct: points to the numerical answers 15, 89.60, and 79.48.
- (b) Not answered: points to the empty box for the 95% CI upper limit.
- (c) Incorrect: points to the 'higher' selection in the dropdown menu.

[Submit your answers](#)

Figure 3

Screenshot of the initial report for the key skill e-assessment example.

Report

FEEDBACK
RETRY

Question (bdm_2_3) MEANS
Your answer, 15, for sample size N is correct.
Your answer, 89.60, for sample mean is correct.
Your answer, 79.48, for lower CI limit is correct.
You did not supply an answer for upper CI limit.
Your answer for mean comparison, [higher], is incorrect.
For this question you scored 3 out of 5.

Your result in total.
You scored 3 out of 5
This gives you a percentage score of 60%.

FEEDBACK
RETRY

Figure 3 demonstrates the initial report, which briefly summarises the students attempt and provides the e-assessment percentage score. Figure 4 displays the ‘Solution’ and the ‘Report’ sections from the in-depth feedback of the student’s attempt at this e-assessment. The ‘Solution’ section supplies the correct answers to the questions that were presented to the student. A pictorial representation of the confidence interval is included along with accompanying text in blue that provides extra feedback information with the aim of aiding students’ understanding of the correct inference to the question posed. The ‘Report’ section indicates, with colour coded marking, what the student has answered correctly (green) or incorrectly (red). It can be seen that the student has one wrong answer and a not answered (NotAns) input.

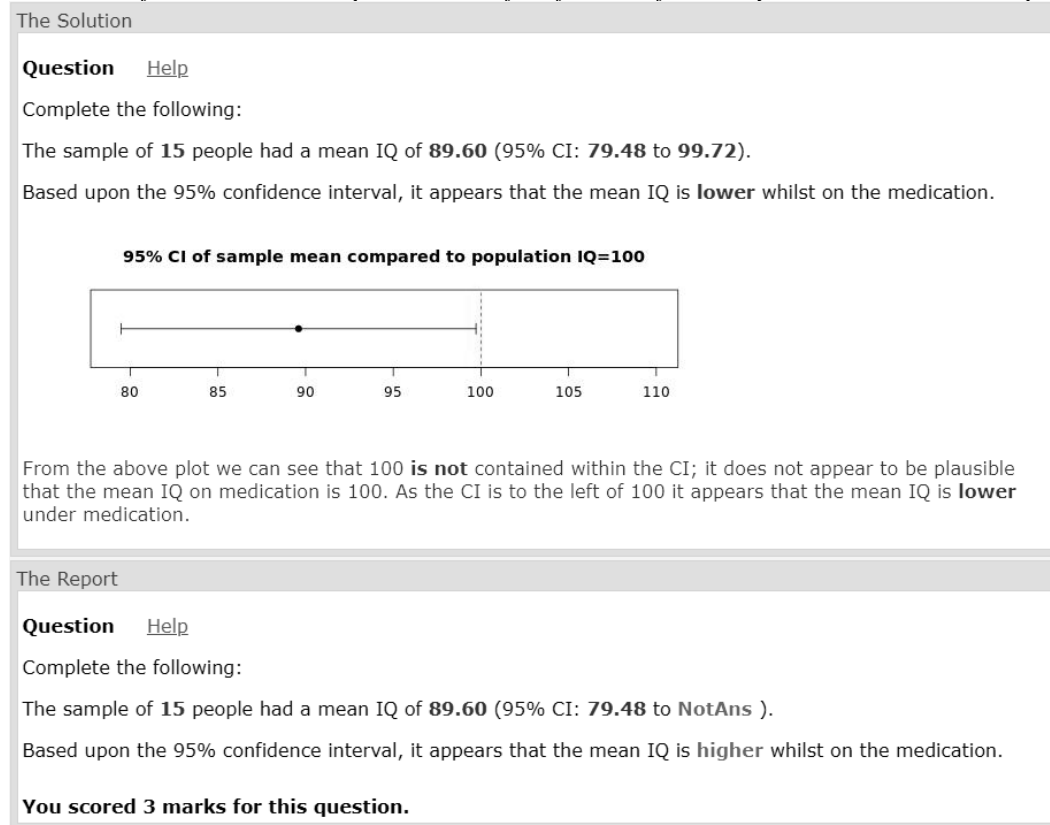
In-class e-exams

Overview

The assessment of the module takes the form of in-class e-exams, which are e-assessments sat under exam conditions, worth 50% of the module mark and a 1,200 word report which students complete in their own time, also worth 50%. There are three in-class e-exams which are equally spaced throughout the 12 week teaching delivery period and assess students’ statistical competency. The final mark that is assigned to the student for the in-class e-exams component, comprises the average of the best two marks from the three e-exams. For each e-exam, a student receives their own unique data set to work on and performs a complete statistical analysis of it using SPSS. To make the scenarios more relevant to the student cohort, both the data and questions asked are related to marketing.

Figure 4

Screenshot of the solution and report sections of the feedback for the key skill e-assessment example.



The first in-class e-exam relates to the data dependent choice of the application of either the one-sample t-test or Wilcoxon signed rank test to evaluate a supplier's claimed average sodium content of mineral bottle water. Data of random sample size is supplied each time the test is accessed, together with a supplier's claimed average that is also random. Furthermore, the data is randomly generated so that with equal chance students would experience both tests and significant or nonsignificant results. The basic statistical analysis tasks tested are the ability to transfer data from CSV format to SPSS, perform an exploratory data analysis for summary statistics, graphics and assumption testing; identification of appropriate statistical test (parametric or nonparametric equivalent), interpretation and reporting of test output. There are, in total, 22 entries (a mixture of numerical and dropdown) and each entry is worth one mark if answered correctly. There are a further three marks available for the correct choice of test to report; this weighting acknowledges the complexity of a decision based upon taught guidelines that consider sample size, skewness and normality. The first in-class e-exam and the key skills e-assessments that support it are available in a showcase website (Weir et al., 2020).

The second in-class e-exam concerns testing for equality of the average amount of time visitors spent on each of two website designs. Based upon initial exploratory data analysis that includes test assumption checking, the student chooses whether to report the independent samples t-test, Welch's t-test (SPSS independent samples t-test output includes this adaptation for when there is evidence of unequal variances) or the Mann-Whitney U test. Each access to the e-exam results in data of

random sample sizes. The data is generated so that there is an equal chance that one of the three tests listed above is the data dependent correct choice, and an equal chance to whether the test results will be significant or nonsignificant. The statistical tasks required build upon those of the first e-exam and include as extras the interpretation of overlapping confidence intervals and the evaluation of equality of variances. Each of the 21 entries (a mixture of numerical and dropdown) are worth one mark if correctly answered. The extra complexity of following the guidelines in order to correctly select the appropriate test to use is worth four marks. The third in-class e-exam concerns the prediction of the sales value of a particular product from the market potential. Each access to the e-exam generates data from a random number of sales territories with varying strengths of positive correlation between the two variables. Students are required to perform and report a correlation and regression analysis. It is required that they report on the strength of the correlation, assess the fit of the regression model, report on outliers, interpret the gradient coefficient and produce predictions together with associated 95% confidence intervals. Unlike the analyses in the other two e-exams, the analysis in the third e-exam does not require the student to make any data dependent decisions on techniques to apply. Each of the 25 entries are worth one mark if correctly answered.

The content of the material assessed in the e-exams prepares students for the subsequent summative coursework. For this piece of coursework, students are given access to a large questionnaire data set in response to a business question. They are required to find something of interest in this data set in order to demonstrate one of the statistical techniques taught in the module. This exercise introduced generic data analysis issues that were not covered in the lectures, namely:

- Reading in data that wasn't immediately suitable for SPSS;
- Excluding observations;
- Combining question categories.

The Statistics team created PowerPoint slideshows for the students to self-learn how to deal with each of these problems.

Controlled conditions procedure

The in-class e-exams that took place at UWE Bristol ran at the same time and in the same room as for students' timetabled PC sessions. Logistically, this was the only sensible way of running the e-exams under controlled, exam conditions, given the large number of students involved. In 2019-20, there were seven PC sessions for the 365 students, each one held at a different time and day of the week. At UWE Bristol, the largest PC lab holds at most 55 students. The time limit for each in-class e-exam is 30 minutes. Taking into account reasonable adjustment students with 50% extra time, this means that each e-exam comfortably fits into the 50 minutes class-time available. A group for each PC session was set up on the VLE and students were allocated to the group that corresponded to their timetabled PC session. Using adaptive release, access to a formative practice e-test for each group was available immediately after the corresponding PC class held the week before the in-class e-exam. The practice e-test was administered in exactly the same format as the in-class e-exam and remained available to that group until midnight the day before their in-class e-exam was scheduled. During its availability period, students were allowed unlimited attempts at the practice e-test getting different data (and hence analysis outcomes) each time they attempted it together with full feedback. Dewis enforced a strict time limit to get students used to working under exam time pressure. For most students this was 30 minutes. However, Dewis automatically adjusted the time available for those students who were allowed extra time in exams, due to reasonable adjustments.

Using adaptive release, the link for the in-class e-exam was only available for a particular PC group during their scheduled PC session. At the start of the PC session students were instructed to log into the VLE to access this link and to open the SPSS software. An examination password protected the in-class e-exam link and this was different for each PC group. Once all the students were ready to start, this examination password was issued and students were directed to input it and “Start the online exam”. This ensured that only students within the room, and hence under exam conditions, were able to access the e-exam.

During the examination, students are only allowed to access Dewis and SPSS. Dewis displays a persistent grey horizontal banner on the bottom of the page which contains details of the student’s identity. On submission of the e-exam, the colour of this horizontal banner changes to pink. This display feature facilitates the invigilation process for e-examinations run using Dewis (Henderson et al., 2016). It provides an easy visual check of the status of a students’ assessment attempt and allows verification that the student taking the e-exam matches that shown on the screen. On submission, students receive confirmation of their submitted answers (see Figure 5). Students may submit their answers ahead of the 30 minute time limit, but are not allowed to leave the examination room until an invigilator has verified that they have submitted (by checking that the banner is pink). This provides extra security, ensuring that a student is not able to continue with the e-exam outside of the examination room. The in-class e-exams were invigilated by staff from the University’s exam invigilator team with academic staff involved with the module available to help with any technical issues.

For all of the e-exams, each student’s attempt is marked instantly on submission. However, disclosure of the mark to the students is delayed until the end of the week. This allows us time to review and analyse the complete spread of marks once all the students have taken each in-class e-exam. Once we have completed the review process, students are able click on the LTI e-exam link and view their submission to get full, bespoke feedback for their attempt as well as their mark.

The number of students taking the module at partner institutions is considerably smaller than for those taking it at UWE Bristol. This means that each partner institution only has to run each e-exam once, as they have sufficient space to examine the whole cohort in one sitting.

Figure 5

Illustration of the persistent banner (bottom) which contains student details and changes colour on submission. In this case, the student has submitted and has received confirmation of their submitted answers. Details of marks and feedback are deliberately withheld at this stage.

The screenshot shows the 'DEWIS @ UWE' online test interface. At the top, the UWE Bristol logo is on the left and 'DEWIS @ UWE' is on the right. Below this, the title 'BDM ONLINE TEST 1' is displayed on the left and 'Submitted' on the right. A prominent red banner reads: 'IMPORTANT: DO NOT CLOSE THIS PAGE UNTIL AN INVIGILATOR HAS SEEN IT.' Below the banner, a grey box contains the message: 'Your submission has been recorded. Thank You. A summary of your submission is given below.' Underneath this is a 'Summary' section with a list of answers: 'Your answer for sample mean is 35.543.', 'Your answer for lower CI limit is 35.449.', 'Your answer for upper CI limit is 35.637.', 'Your answer for mean comparison was 'No the average sodium content appears to be a higher value'', 'Your answer for median is 35.600.', 'Your answer for median comparison was 'higher than'', and 'Your answer for skewness statistic is 0.613'. At the bottom, a grey banner shows 'BDM ONLINE TEST 1' on the left and 'Username: r19ya' on the right.

Evaluation

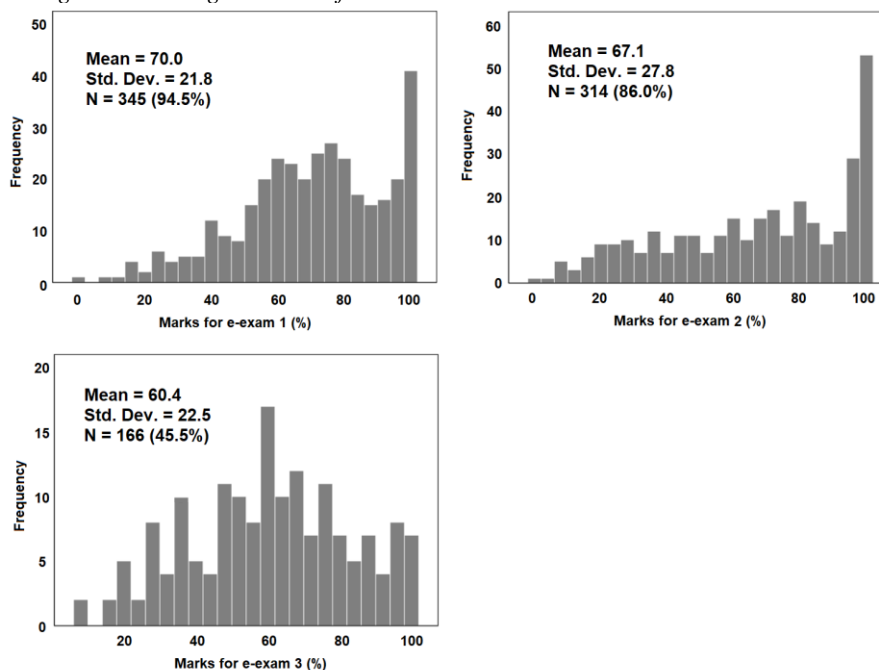
Student Performance

The module has now run for three years and student results for these three academic years are summarised in Table 1. The in-class e-exam mark average and pass rate are similar each year; the main difference across the years being an increase in student numbers. It is not appropriate to compare with previous student cohorts as BDM is a new module, created as a result of a curriculum refresh process and the statistical material taught within it was not previously covered at this level. Given the challenging nature of the material taught, it is very encouraging to note that the average mark and pass rates for BDM is at least comparable to the students' performances in their other modules.

Table 1*Summary of student results for all academic years to date.*

Academic Year	Number of students		Overall e-exam average (%)	In-class mark	Overall e-exam pass rate (%)
	Active on the module	Sat at least one in-class e-exam			
2019-20	365	356	69.4		88.2
2018-19	316	306	74.6		91.5
2017-18	276	270	71.7		87.8

The analysis presented in the rest of this section relates specifically to the 2019-20 academic year. The breakdown of the results for each e-exam is shown in Figure 6. The first e-exam was sat by 345 students and the majority of these did very well, with over 40 students scoring 100%. The second e-exam was taken by less students; the results are more spread out but the cohort's performance is again strong. The decline in the number of students attempting the third e-exam is due to the final mark for the in-class e-exam component of assessment being the best two from three; many students with high marks from the first two tests elected not to take the final e-exam. This is reflected in the lower mean mark for the third e-exam. A total of 115 students took all three e-exams.

Figure 6*Histograms showing the marks for the three e-exams.*

Of those students that sat each e-exam, approximately half had previously tried the corresponding practice e-test (e-exam 1: 48%; e-exam 2: 55%; e-exam 3: 49%). Histograms showing the number of (non-zero) attempts on each practice e-test are presented in Figure 7. Typically students made on average 4-5 attempts at each practice e-test (e-exam 1: $Mdn = 5$, $IQR = 2 - 9$; e-exam 2: $Mdn = 5$, $IQR = 3 - 10$; e-exam 3: $Mdn = 4$, $IQR = 2 - 6$). For each e-exam, those students who attempted the

corresponding practice e-test at least once experienced on average a significant increase in their e-exam mark compared to those students who did not practice (e-exam 1 increase 24.5 (95% CI 20.6 – 28.3) marks out of 100; e-exam 2: 33.4 (95% CI 28.4 – 38.4); e-exam 3: 22.1 (95% CI 16.0 – 28.2)).

There are 23 key skills e-assessments employed on the module; a histogram of the total number of attempts made by students on these is shown in Figure 8. On average each student made 31 attempts, which equates to over 11,000 assessment attempts on the module. The correlations of the marks for each of the three e-exams with the highest scores from their respective practice e-tests and key skills e-assessments are shown in Table 2. All are significant, positive correlations with strengths ranging from moderate to weak. The key skills correlations are almost exclusively of lesser size than for the practice e-tests which reflects that they contribute to just a part of the e-exam. Note that only 11 of the key skills e-assessments contain material that is directly examined in the e-exams.

A high number of students viewed their electronic feedback (e-exam 1: 81%; e-exam 2: 82%; e-exam 3: 54%). This is much higher than the 10% that is typical for paper-based work at UWE Bristol. One possible reason for the large number of students viewing their feedback is that they are familiar with and used to the process of receiving feedback from the formative e-assessments. Another possible reason is that the feedback for the e-exams is made available much sooner than is the case for paper-based work, which at UWE Bristol is typically four weeks. Research (Race, 2014) shows that feedback has to be quick to be effective, while students still remember clearly the work they were engaged in and online exams are one way of achieving this.

Monitoring of Student Engagement

Attendance at classes is not compulsory at UWE Bristol and as such, no formal monitoring of student attendance takes place. However, through the Dewis management system, staff can view the marks and feedback of each and every student attempt. In addition, Dewis automatically updates the University's VLE with the highest scoring attempt from each of the key skill e-assessments as well as e-exam results for each student. By looking at these marks, or lack of them, in the Grade Centre the module team is able to identify non-engagers and strugglers easily and hence target them appropriately. For example, struggling students will be those with low scores and such students can be recommended to access further support such as attending espressoMaths, which is the University's drop-by mathematical support service (Henderson & Swift, 2011). Those students who did not attempt the first e-exam were contacted by email and were reminded of the support available for them. It is encouraging to see from Table 1, that only a small number of students missed all three e-exams.

Figure 8

Histogram showing the number of attempts made on the key skills e-assessments.

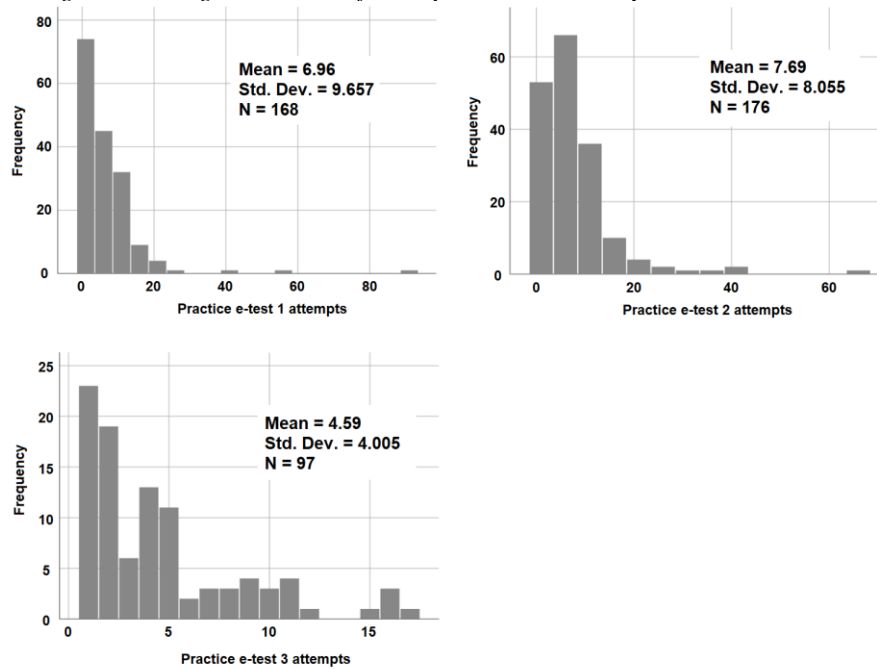


Figure 7. Histograms showing the number of attempts by those engaging with each practice e-test.

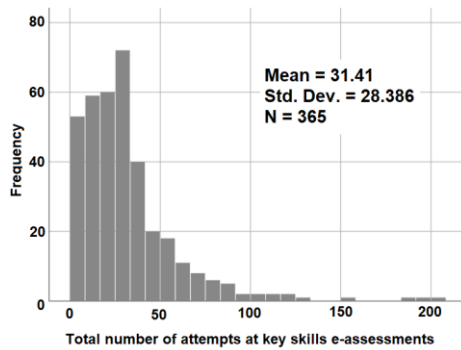


Table 2

Correlations of marks for the three e-exams with the marks for each respective practice e-test and key skills e-assessment.

	E-exam 1	E-exam 2	E-exam 3
	Pearson Correlation, (Sig. (2-tailed)), N		
Practice e-tests			
E-test 1	.658 (.000) 161		
E-test 2		.644 (.000) 165	
E-test 3			.483 (.000) 76
Key skills e-assessments			
Means and 95% confidence intervals	.237 (.000) 260		
Skewness and kurtosis	.313 (.000) 235	.277 (.000) 215	
Shapiro-Wilk normality test	.218 (.012) 201	.248 (.001) 184	
One sample t-test	.330 (.000) 279		
Wilcoxon signed rank test	.260 (.000) 238		
Which one-sample test?	.199 (.005) 197		
Independent t-test		.507 (.000) 234	
Mann-Whitney U test		.389 (.000) 190	
Which independent samples test?		.436 (.000) 157	
Pearson's correlation			.372 (.000) 119
Simple linear regression			.500 (.000) 112

Student Feedback

Feedback from students, via module evaluations, shows that most students like the in-class e-exams, recognising that “it is fair for all students” because it is not possible to “cheat” as is the case for coursework. In the first year of operation (2017-18) some students commented that there was not enough linkage between the statistics and business strands of the module. In response to this, from 2018 onwards, there has been greater collaboration between these two strands with, for example the business tutorials using and reinforcing the relevant statistical topics. Another strengthening of the linkage related to improving the statistical support for the business coursework. This was achieved by creating extra PowerPoint slideshows targeted at helping students with generic data analysis issues with the business report data that were not covered in the statistics lectures.

Ongoing use of SPSS

The Business School students build upon their statistical knowledge in BDM through further exposure to statistics in their second year. These students are required to attempt coursework that involves analysis of real data in a substantial research project. For this coursework students may choose to analyse either quantitative or qualitative data and are not required to use SPSS. Anecdotal evidence from academics teaching and assessing these second year students is that those students who do choose to use SPSS to analyse quantitative data display a high degree of statistical competency. Academics have also reported that they have also seen an increase in the use of SPSS as opposed to Excel. The comprehensive suite of materials developed within BDM, including the portfolio of completed templates, is available to students even after they have completed this module. Students are thus able to refer back to them when they are required to conduct their own statistical report writing and facilitates their transition to becoming independent researchers.

Discussion

We have presented a teaching and assessment approach that supports students to learn how to use SPSS to analyse data. The approach is centred on using innovative Dewis e-assessments to generate random data and to automatically mark submissions that match SPSS output. The use of formative key skills e-assessments with in-built Help pages enables students to independently learn the mechanics of SPSS output generation and the subsequent interpretation of it. With repeated practice, students can use the feedback that is provided by Dewis immediately on submission to check their results and understanding. E-exams are used to assess these skills, which are administered in class under exam conditions. Extra security measures have been introduced into Dewis in order to make this process robust and straightforward for non-specialists to invigilate. The fact that the formative and summative e-assessments are marked immediately by Dewis represents a huge saving in marking time compared to human marking and allows students to have instant feedback on their work. There are initial set up costs; typically a new e-assessment can be designed, and tested within a day by one person. However, once established these e-assessments may be used with only minor edits required each year. Student results have been excellent with high marks and pass rates being achieved on the in-class e-exam component each year. The fact that the final mark is calculated as the best two from three e-exams does however mean that many students choose not to sit the third e-exam. Students who attempt the practice e-tests and formative key e-assessments perform better in the e-exams. Feedback from module evaluations shows that most students like the in-class e-exams. There is some evidence that the use of SPSS by students in their subsequent years of study is increasing.

A similar approach, to that described in this paper, has been used on a module delivered to management students at UWE Bristol. This module has approximately 400 students and has the same assessment profile. The main difference is that there is less dovetailing between the two parts of the module (Statistics and Business). For this module, the third in-class e-exam in the 2019-20 academic year was run remotely due to Covid-19. It was scheduled to be held the week after the UK went into lockdown for the first time (Stewart et al., 2020). Because of the change in working practices, students were required to access the e-exam remotely on the day and at the time that it would have happened if on campus. No additional exam password was required, but students had the same time constraints for submission as for the previous two e-exams. There were no technical issues and all students who started the e-exam submitted successfully. Students and staff were impressed with how seamlessly the delivery transitioned to online. A similar statistics e-assessment approach has also been implemented for a module delivered to environmental scientists and is being brought into our new curriculum in Mathematics from 2020 onwards. It is hoped that the work presented in this paper will generate interest from the wider statistics community. Further e-assessment examples and contact details can be found at the Dewis-R Statistics Resources website (Weir, 2020) for those interested in this new approach.

Conclusions

The e-assessment approach described in this case study has enabled a large cohort of students to independently learn SPSS and to attain the required statistical competency and understanding to be successful on the module. We have found that students have engaged well with the formative e-assessments, which are marked instantly and provide comprehensive feedback. Repeated use of these formative e-assessments allows students to independently practice and check their understanding. About half of the students passed the e-exams without utilising the practice e-tests beforehand, but those that did practice achieved significantly higher marks. Although this case study has focused on one particular module delivered to business students a similar e-assessment

approach. Covid-19 has forced many higher education institutions to move to online or blended deliveries and increased usage of e-assessment is one way of achieving this.

References

- Aljandali, A. (2016). *Quantitative analysis and IBM® SPSS® statistics*. Springer.
<http://doi.org/10.1007/978-3-319-45528-0>
- Blanco, M., & Ginovart, M. (2010). Moodle quizzes for assessing statistical topics in engineering studies. *Proceedings of the Joint International IGIP, SEFI Annual Conference*. 19th - 22nd September 2010, Trnava, Slovakia.
- Bothwell, E. (2020). UK universities favour blended learning approach for 2020-21. *The Times Higher Education*. <https://www.timeshighereducation.com/news/uk-universities-favour-blended-learning-approach-2020-21>
- Brink, R., & Lautenbach, G. (2011). Electronic assessment in higher education. *Educational Studies*, 37(5), 503-512. <http://doi.org/10.1080/03055698.2010.539733>
- Collins, C., Ripley, M., & Roads, M. (2003). The opportunities for ICT in assessment. In C Richardson (ed.), *Whither Assessment?* Qualifications and Curriculum Authority.
- Delmas, R., Garfield, J., Ooms, A., & Chance B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, 6(2), 28–58. [https://iase-web.org/documents/SERJ/SERJ6\(2\)_delMas.pdf?1402525007](https://iase-web.org/documents/SERJ/SERJ6(2)_delMas.pdf?1402525007)
- Dewis Development Team. (2012). *Dewis welcome page*. <http://dewis.uwe.ac.uk>
- Garfield, J. (2006). *ARTIST web site*. <https://apps3.cehd.umn.edu/artist>
- Gray, L., Sheppard, M., & Ferrell, G. (2016). *Jisc Project: Electronic management of assessment*. <https://www.jisc.ac.uk/rd/projects/electronic-management-of-assessment>
- Greenhow, M., & Gill, M. (2008). How effective is feedback in computer-aided assessments? *Learning, Media and Technology*, 33(3), 207-220.
<http://doi.org/10.1080/17439880802324145>
- Gwynllwy, R., & Henderson, K. (2009). DEWIS: a computer aided assessment system for mathematics and statistics. *CETL-MSOR 2008 Conference Proceedings*, 38-44.
http://www.mathstore.ac.uk/repository/doc.php?doc=CETLMSOR2008_Proceedings.pdf
- Gwynllwy, R., Weir, I., & Henderson, K. (2015). Using DEWIS and R for multi-staged statistics e-assessments. *Teaching Mathematics and its Applications*, 35(1), 14-26.
<http://doi.org/10.1093/teamat/hrv018>
- Henderson, K., Gwynllwy, R., & Hooper, A. (2016). Using electronic exams to provide engineering mathematics students with rapid feedback. In B. Alpers, U. Dinger, T. Gustafsson & D. Velichova (Eds.), *18th SEFI Mathematics Working Group Seminar on Mathematics in Engineering Education Proceedings* (pp. 105-111). SEFI Eur. Soc. for Engineering Education.
- Henderson, K., & Swift, T. (2011). espressoMaths: A drop-by station. *MSOR Connections*, 11(2), 10-13.
- Hodgson, P., & Pang, M. (2012). Effective formative e-assessment of student learning: a study on a statistics course. *Assessment & Evaluation in Higher Education*, 37(2), 215–225.
<http://doi.org/10.1080/02602938.2010.523818>
- Horton, N. (2015). Challenges and opportunities for statistics and statistical education: Looking back, looking forward. *The American Statistician*, 69(2), 318-145.
<http://doi.org/10.1080/00031305.2015.1032435>
- Idris, K. (2018). Teaching and learning statistics in college: how learning materials should be designed. *Journal of Physics Conference Series*, 1088, 012032.
<https://iopscience.iop.org/article/10.1088/1742-6596/1088/1/012032>
- IBM Corp. (2019). IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp.
- Journal of Statistical Education. (2020). An official journal of the American Statistical Association. *Taylor & Francis*. <https://www.tandfonline.com/toc/ujse20/current>

- Kuikkaa, M., Kitolab, M., & Laaksob, M-J. (2014). Challenges when introducing electronic exam. *Research in Learning Technology*, 22, 22817. <http://doi.org/10.3402/rlt.v22.22817>
- Mustafa, R. (1996). The challenge of teaching statistics to non-specialists. *Journal of Statistical Education*, 4(1). <http://doi.org/10.1080/10691898.1996.11910504>
- R Core Team. (2018). *R: A language and environment for statistical computing*, R foundation for statistical computing. <http://www.R-project.org/>
- Race, P. (2014). *Making learning happen*. Sage Publications.
- Roads, M. (2016). *JISC e-assessment survey. Final report*. http://repository.jisc.ac.uk/6278/14/e-Assessment_survey_report_May_2016_FINAL.pdf
- Sangwin, C. (2013). *Computer aided assessment of mathematics*. Oxford University Press.
- Sikurajapathi, I., Henderson, K., & Gwynllyw, R. (2020). Using e-assessment to address mathematical misconceptions in engineering students. *International Journal of Information and Education Technology*, 10(5), 356-361. <http://doi.org/10.18178/ijiet.2020.10.5.1389>
- Simpson, J. (1995). Teaching statistics to non-specialists. *Statistics in Medicine*, 14(2), 199-208. <http://doi.org/10.1002/sim.4780140210>
- STEM Learning. (2009). Relevant and engaging statistics. *Royal Statistical Society Centre for Statistical Education*. <https://www.stem.org.uk/cx5ba>
- Stewart, H., Mason, R., & Dodd, V. (2020, March 23). Boris Johnson orders UK lockdown to be enforced by police. *The Guardian*. <https://www.theguardian.com/world/2020/mar/23/boris-johnson-orders-uk-lockdown-to-be-enforced-by-police>
- Weir, I., (2020). Dewis-R Statistics resources website. <http://dewisprod.uwe.ac.uk/resources/stats/index.html>
- Weir, I., & Gwynllyw, R. (2017). Using the Dewis-R interface to create statistics e-Assessments, IATED (ed.). *Edulearn17 Proceedings*. <http://doi.org/10.21125/edulearn.2017.2385>
- Weir, I., Gwynllyw, R., & Henderson, K. (2015). Using technology to inspire and enhance the learning of statistics in a large cohort of diverse ability in IATED (ed.). *Edulearn15 Proceedings*.
- Weir, I., Gwynllyw, R., & Henderson, K. (2020). Business decision making: In-class test 1. <http://dewisprod.uwe.ac.uk/resources/bdm/ict1/>
- Willis, R., Yeigh, T., Lynch, D., Smith, R., Provost, S., Turner, D., & Sell, K. (2018). *Towards a strategic blend in education: a review of the blended learning literature*. Oxford Global Press.
- Wood, B., Mocko, M., Everson, M., Horton, N., & Velleman, P. (2018). Updated guidelines, updated curriculum: The *GAISE College Report* and introductory statistics for the modern student. *CHANCE*, 31(2), 53-59. <http://doi.org/10.1080/09332480.2018.1467642>
- World Health Organization. (2020). *Coronavirus disease (COVID-19) Pandemic*. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>

Appendix 1

Example of part of a pre-prepared Word template

Week 4: Wilcoxon signed rank test

The voltages of a random sample of nine 1.5 volts batteries were tested with the following results:

1.47 1.49 1.51 1.50 1.46 1.52 1.47 1.48 1.50

Test to see if on average the batteries are 1.5 volts.

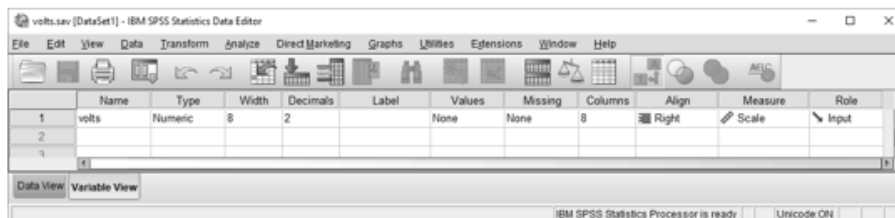
As we have a small sample ($N < 10$) we will be using the nonparametric test to see if the average voltage is 1.5 volts (i.e. one-sample Wilcoxon signed rank test).

However, we still *Explore* the data though to get to know it better; you still need to screen your data for potential outliers and you still would need to be able to report numerical and graphical summary statistics of the data! However we will not be concerned with the output relating to normality.

INSTRUCTIONS

You will need to type the data into SPSS. Complete the **Variable View** as below.

- Note in the CSV file the precision the data is recorded to and set to 2 in **Decimals**.
- The variable is *Scale Measure*.



Then create the required SPSS output and paste where prompted.

In the text below *** prompts where you have to type a numerical value read from SPSS output.

Text in bold indicates where you have to choose the appropriate comment based upon supporting output e.g. we **do/do not reject** H_0 .

Exploratory Data Analysis

PASTE BOXPLOT HERE

The boxplot gives the impression that the average voltage is **lower/higher** than 1.5 volts.

The boxplot reveals there are *** outliers.

PASTE DESCRIPTIVES TABLE HERE