Engaging University Students in Statistics through Innovative Approaches – with examples from Africa

Roger Stern\textsuperscript{1}, Danny Parsons\textsuperscript{2} and James Musyoka\textsuperscript{3}

\textsuperscript{1}University of Reading & Statistics for Sustainable Development, UK
\textsuperscript{2}IDEMS International, UK
\textsuperscript{3}Maseno University, Kenya
Contents

- A statistics problem solving course for MSc Mathematical Science students
- Innovations from the course
- Might STACK be a “game-changer”?
- Innovations in teaching at Maseno University, Kenya
MSc Mathematical Science

Aim to prepare students with skills to solve problems that can contribute to the development of Africa

“Statistics Problem Solving” course as first in stats stream supported by RSS

Parallel of mathematical problem solving at AIMS
Student Diversity & Background (2019)

- 47 MSc students from 12 African countries
- From no statistics courses to full statistics degrees
- Theoretical statistics teaching – analysis focused
- Relatively little use of data/software
Statistics?

Data ownership → Planning data flow → Planning data collection

Quality assurance

Data collection → Data entry and organisation → Statistical analysis

Data storage and access → Dissemination & feedback to data originators

Quality assessment

Interpretation and write-up
Aims and Principles

Students should:

- understand the role of statistics in solving real problems with data
- be comfortable with producing and interpreting descriptive tools
- understand the broader processes within statistics from design and collection to interpretation and presentation of results

Support students to:

- engage in statistics through experiential learning
- use real world problem and research related projects that have genuine complexity
  - without single “correct” answers

And

- Be appropriate for students with very varied previous knowledge
MSc Course structure

- Three week course - 15 two-hour sessions
- There are 4-6 permanent AIMS tutors supporting the students
  - Tutorial sessions each week are led by tutors
- One piece of assessed work each week – tutors support the marking
- Plus quizzes
  - We introduced electronic quizzes in our problem-solving course
  - This involved repeatable mastery + single test per week
- (Group) project + presentation at the end of the course
Statistics Problem Solving Course

Week 1:
- What is statistics? RSS definition: design, collection, organisation, interpretation and communication
- What does a statistician do? Example consultancy questions - mostly about design.
- Statistical “games” simulating an experiment and a survey
- Assessed work is a report, with “Introduction, methods, results, conclusions” – like a paper

Week 2:
- Working with data in a spreadsheet and a statistics package & interpreting results
- Randomised, simulated data from large (digital) version of experimental “game”.
- A good analysis explains variability – ANOVA as a descriptive tool
- A good analysis relates to the objectives of the study
  • (statistics students were used to significance tests)

Week 3:
- Mini-projects in groups (3 or 4) with a final report and presentation
Statistical Games

**Essentially simulations**:  
- Design the study, then collect the data, enter, then analyse and report  
- No substitute for real data, but possible within (say) 2 hours

**Experimental games**:  
- Design tomato yield experiment  
- Eight treatments – 2 by 2 by 2 factorial in 2 blocks, each of size 6!  
- Repeated for 2 years – 24 plots overall  
- Design deliberately not obvious!  
- Larger version (week 2 of the course) is usually 60 times larger

**Survey game**:  
- Crop-cutting survey – loosely based on rice survey in Sri Lanka  
- 10 villages of different sizes, with total of 160 farmers  
- So multilevel data, with village, farmer, plot – and multiple objectives
Visualisations from larger experimental game data

1,440 observations
Interactive questions to test understanding

- The course was highly practical
- Students also wanted & needed some theory of basic concepts
- Quiz questions tested understanding of concepts – not formulas
- The idea of “estimating” and not “calculating” was alien to some
- Randomised questions allowed for mastery
- Quizzes to help you learn and not to test you, took some getting used to
What is standard deviation?

Source: CAST Computer-Assisted Statistics Textbooks, Doug Stirling
http://cast.massey.ac.nz/
Random datasets

Below are stacked Dot Plots and Box Plots for four different sets of data. Which box plot matches with each dot plot?

Dot plots:

D1

D2

D3

D4

Box Plots:

A

B

C

D
Week 3: Mini-projects using real world data

- Groups of 3 or 4 for each mini-project
- One project only, but ideas shared through the week, so some learning of each topic
- Analyses used a mixture of R, R-Instat, Python (a few) and/or Open Office
- Support by lecturers and tutors assigned to groups
- Final day was a presentation by each group. They also submitted a written group report

*Example topic – more on next slide:*
  - Week 1: Simple (paper) survey of students’ previous experience and views on statistics
  - Week 2: Same questions administer by tutors and collected using ODK (Open Data Kit)
  - Week 3: One topic was for the group to design, collect and analyse their own ODK survey
Week 3 – Mini project topics

- **Analysing corruption** “red flags” in public procurement.
  - Using open World Bank data (200,000 records from over 140 countries.)

- A 2017 **on-farm trial** of low-cost fertilizers
  - Involving 1,700 mainly women farmers in Niger

- Cameroon **climatic data analysis** using daily data from their Met Service.

- Analysing **monitoring and evaluation data** from PICSA in Ghana
  - PICSA = A participatory approach to climate service with small-scale farmers

- **IFAD poverty survey data** of 1,300 respondents and over 400 variables from 2018 in Lesotho.

- Analysis of **Nutrition data** – data from West Africa

- Analysis of **student data from electronic assessment quizzes**

- **Tidy data**. Based on a paper by Hadley Wickham
  - Including his data from Mexico (500,000 cases) plus messy Cameroon climatic data.
An eye-opening course:
- 38/47 said their view of statistics had changed because of the course
- “I now know that statistics is more practical than theoretical as taught in schools”
- “Computers can make calculations, but it is for the statistician to give to those calculations a real signification.”

Motivating & simulating an interest in statistics:
- “this course gave me the interest in statistics, I didn't know before that statistic is a very good field”
- “statistics has never been more fun”

Exposure to the role of statistics in solving real problems with data:
- 40 out of 47 students said the course prepared them “a lot” to use statistics to solve problems
- “I now understand what a typical statistician should do”,
- “now I find that it is a promising field that will help Africa to solve several problems”
- “it is useful in solving the problems of society.”

The broader subject:
- “statistics is not all about Data Analysis, but it starts even before the problem, data collection until the final conclusion”
STACK (System for Teaching and Assessment using a Computer algebra Kernel)

- System for automated assessment of mathematics
- Functionality of a computer algebra system (Maxima)
- Beyond multiple choice – mathematical expressions evaluated by computer algebra

- Works in Moodle (course management system)
- Used in hundreds of sites in multiple languages
- Free and open source (https://github.com/maths/moodle-qtype_stack)
Features of STACK

- Algebraic input
- Establish mathematical properties (infinite correct answers)
- Intelligent feedback
- Randomisation within questions
- Exact solutions
- Line by line working
The precipitation in 63 Eastern African cities in August was measured and the results were grouped in intervals of 6mm. The results are displayed in the graph below. Estimate ‘by eye’ the values of the mean and median precipitation.

Incorrect answer.
Your relative positioning of the mean and median do not reflect the skewness in the data. Remember that outliers affect the mean but not the median, hence with skew data the mean is ‘pulled’ out towards the extreme values while the median is not affected by individual extreme values.

Correct answer, well done.
Your answer for the mean is close enough. The actual value is 29 mm.

Incorrect answer.
You are a bit too far from the actual median value of 21 mm. Your estimate of the median is more than 20% away, with practice you should be able to do better.
Specific Feedback & Intelligent marking

Here is an incomplete analysis of variance (ANOVA) table for the tomato data for **2 years together**.

Complete the ANOVA table.

Give the values to 1 decimal place.

<table>
<thead>
<tr>
<th>Term</th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light</td>
<td>1</td>
<td>9.5</td>
<td>9.5</td>
<td>0.8</td>
</tr>
<tr>
<td>Heat</td>
<td>1</td>
<td>46.1</td>
<td>46.1</td>
<td>4.0</td>
</tr>
<tr>
<td>Variety</td>
<td>1</td>
<td>✔ Well done.</td>
<td>✔ Well done.</td>
<td>✔ Well done.</td>
</tr>
</tbody>
</table>

Total 23 186.1

- **Light**
  - Correct F value based on your mean sum of squares and residual sum of squares.
- **Heat**
  - Correct F value based on your mean sum of squares and residual sum of squares.
- **Variety**
  - Correct mean square value based on your sum of squares and Df.
- **Residuals**
  - Incorrect answer.
  - Residual is the "left over" from the variables. So the residual degrees of freedom is the total degree of freedom of the data minus the degrees of freedoms of the variables.
  - Correct residual mean square value based on your residual sum of squares and Df.
Interactive Graphics

Move the quantiles and median lines in the box plot so that it is representative of the data displayed in the histogram below.
STACK for statistics

- STACK is widely used for mathematics courses
- Less used for statistics
- Innovation to use random datasets with JSX graphics in STACK
- Begin to also include interactive elements
Find a vector $\mathbf{v}_4$ so that together the four vectors
\[
\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{v}_4 \right\}
\]
do not form a basis of $\mathbb{R}^4$. 
Consider the real function

\[ f(x) := \begin{cases} 
1 & \text{for } x \leq -2, \\
p(x) & \text{for } -2 < x < 0, \\
\sin(x) & \text{for } 0 \leq x.
\end{cases} \]

Find the cubic polynomial \( p(x) \) which makes \( f(x) \) continuously differentiable.

\[ p(x) = x^2(x+3)/4 \]
Your answer is partially correct.
Your answer does not satisfy $p(-2) = 1$.
Your answer does not satisfy $p'(-2) = 0$.
Your answer does not satisfy $p'(0) = 1$.

Your answer is plotted below, although part of your graph might appear out of range of the plot!

Marks for this submission: 0.25/1.00.
Maseno University, Kenya
School of Mathematics, Statistics and Actuarial Science

First year courses

- Calculus I (350 students)
- Introduction to Linear Algebra (700 students)
- Descriptive Statistics (1,000 students)
- Introduction to Probability theory (350 students)

- One lecturer per course – no teaching assistants
- So no one wants the first year courses!
Challenges with Assessment

**Lecturers**
- Limited time for marking – can’t provide good feedback
- Don’t have time for other activities (projects/research)

**Students**
- Lack of high quality feedback
- Little interaction with content outside lectures
Assessment strategies

Electronic Assessments
- Accessible using a smart phone

Mastery Quizzes
- Formative assessment with immediate feedback
- To enhance student understanding
- Unlimited attempts and access

Test Quizzes
- Summative assessment
- To differentiate students
- Limited attempts and access
<table>
<thead>
<tr>
<th>Quiz format</th>
<th>Mastery Quizzes</th>
<th>Test Quizzes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attempts allowed</td>
<td>Infinite</td>
<td>1</td>
</tr>
<tr>
<td>Deadline</td>
<td>End of Semester</td>
<td>End of each week</td>
</tr>
<tr>
<td>Feedback</td>
<td>Immediate</td>
<td>End of each week</td>
</tr>
<tr>
<td>Restriction</td>
<td>None</td>
<td>&gt;70% on Mastery Quiz</td>
</tr>
<tr>
<td>Percentage of</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Continuous Assessment</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Most beneficial aspects to students

1. Getting feedback on my quiz answers
2. Multiple attempts at the Master Quizzes
3. Mastery Quizzes remain open for whole semester
4. Having regular weekly quizzes
5. Having regular interactions with other students to discuss STACK questions
6. Randomised questions in the same quiz
7. Being able to interact with students/lecturers in the forums
Challenging aspects for students

1. Accessing internet
2. Registering and using course site
3. Understanding the course content
4. Typing answers to quizzes in STACK
5. Using the quiz system
6. Accessing a device
Estimating standard deviation by eye

The following diagram shows the height in cm of 74 adults in a village.

Estimate the standard deviation of heights

8 cm

Correct answer, well done.

Well done you are within 10% of the actual standard deviation!

Use the following as a guide.

An interval that is 2s wide usually covers about the middle two thirds of the crosses (i.e. $\bar{x} \pm s$).

An interval that is 4s wide usually covers about the middle 95% of the crosses (i.e. $\bar{x} \pm 2s$).

The range of the data is usually between 5s and 6s (i.e. $\bar{x} \pm 2.5s$ to $\bar{x} \pm 3s$).

In this case the mean is about 163 and you should be able to estimate the standard deviation close to the real value of 7.4.
Comments from students

Am strongly supporting this mode of learning because it motivates one to learn so that the content to stick in mind. This helps one to have easy time in doing the quizzes and understanding the content. All the time you have an idea of all the content in each topic of the unit. Also the review of the quizzes is very beneficial. This is because it is directing one to the right way of solving a certain problem. This helps one to identify errors and correct.

I suggest that the questions in mastery quiz should be added to enhance more understanding of the course content. I also suggest that group discussions should be formed which would be part of the quiz to enhance the mastery of the content. I appreciate the lecturer for the excellent work, the quizzes have really helped me.

Encourage other lecturers to do the same

The course boosted my revision towards the course work.

Excellent!! a quite encouraging and beneficial method of taking the course. I wish every unit could be conducted the same way. It has enabled us to know deeper what we are taught in class. Bravo Dr. Michael and your team for a such brilliant initiative.

Two sitting CATS are better than the STACK quizzes.
What’s next & more initiatives

- Broaden RSS statistics courses support to other AIMS centres? (Rwanda, Ghana, South Africa, Senegal)
- Adapt Statistics Problem Solving course for undergraduates
- Develop more open course materials using STACK for universities to use and adapt
- STACK for schools
References


Acknowledgement

Thanks to the Royal Statistical Society for supporting the series of statistics courses delivered at AIMS Cameroon.

Thanks to all the students, tutors, lecturers & staff for embracing these ideas at AIMS & Maseno.