

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

Roger Stern<sup>1</sup>, Danny Parsons<sup>2</sup> and James Musyoka<sup>3</sup>

<sup>1</sup>University of Reading & Statistics for Sustainable Development, UK

<sup>2</sup>IDEMS International, UK

<sup>3</sup>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

# African Institute for Mathematics Sciences (AIMS) Cameroon (Limbe)



## Background

MSc Mathematical Science

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

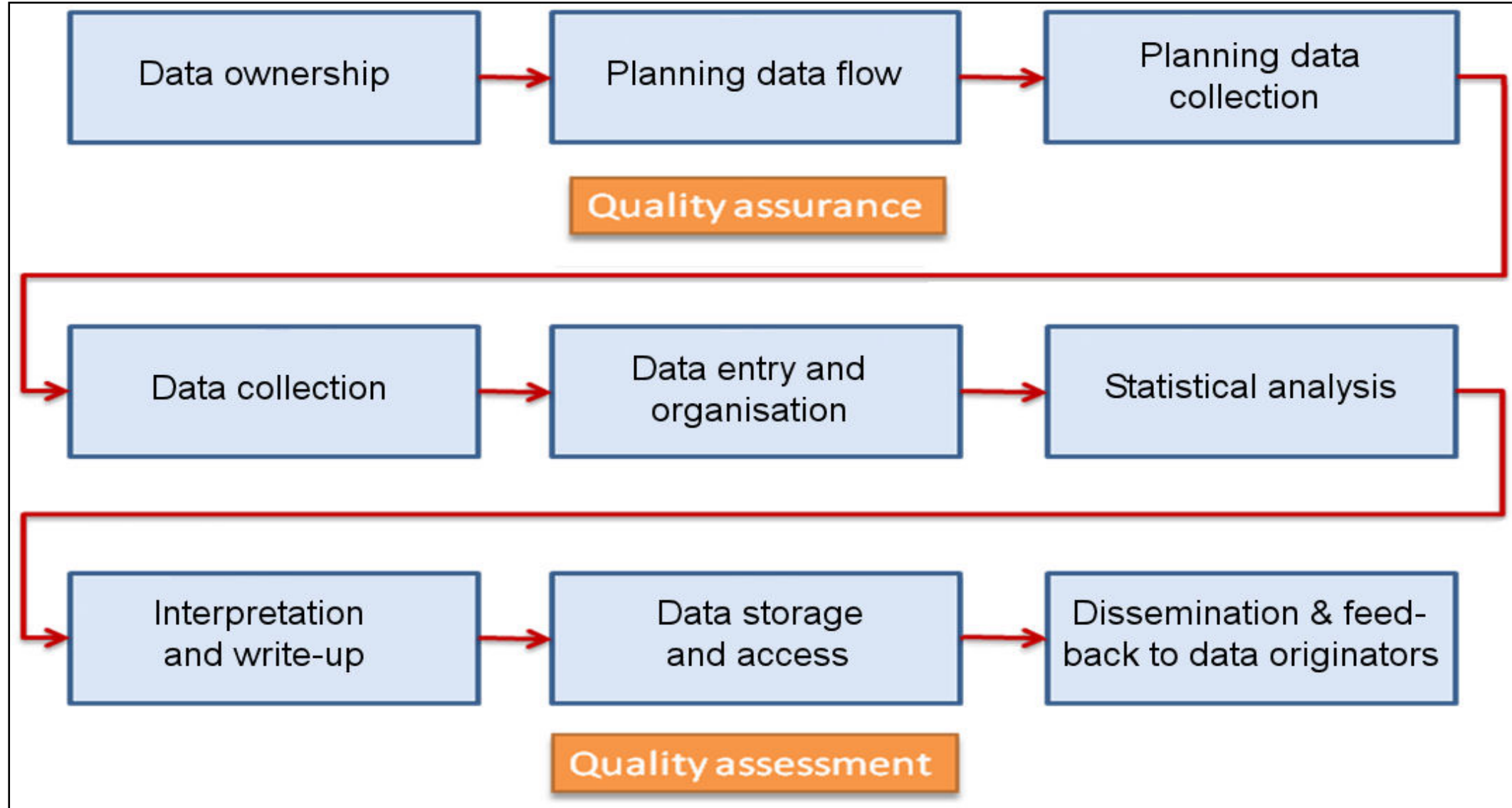
“Statistics Problem Solving” course as first in stats stream supported by RSS

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?



# 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

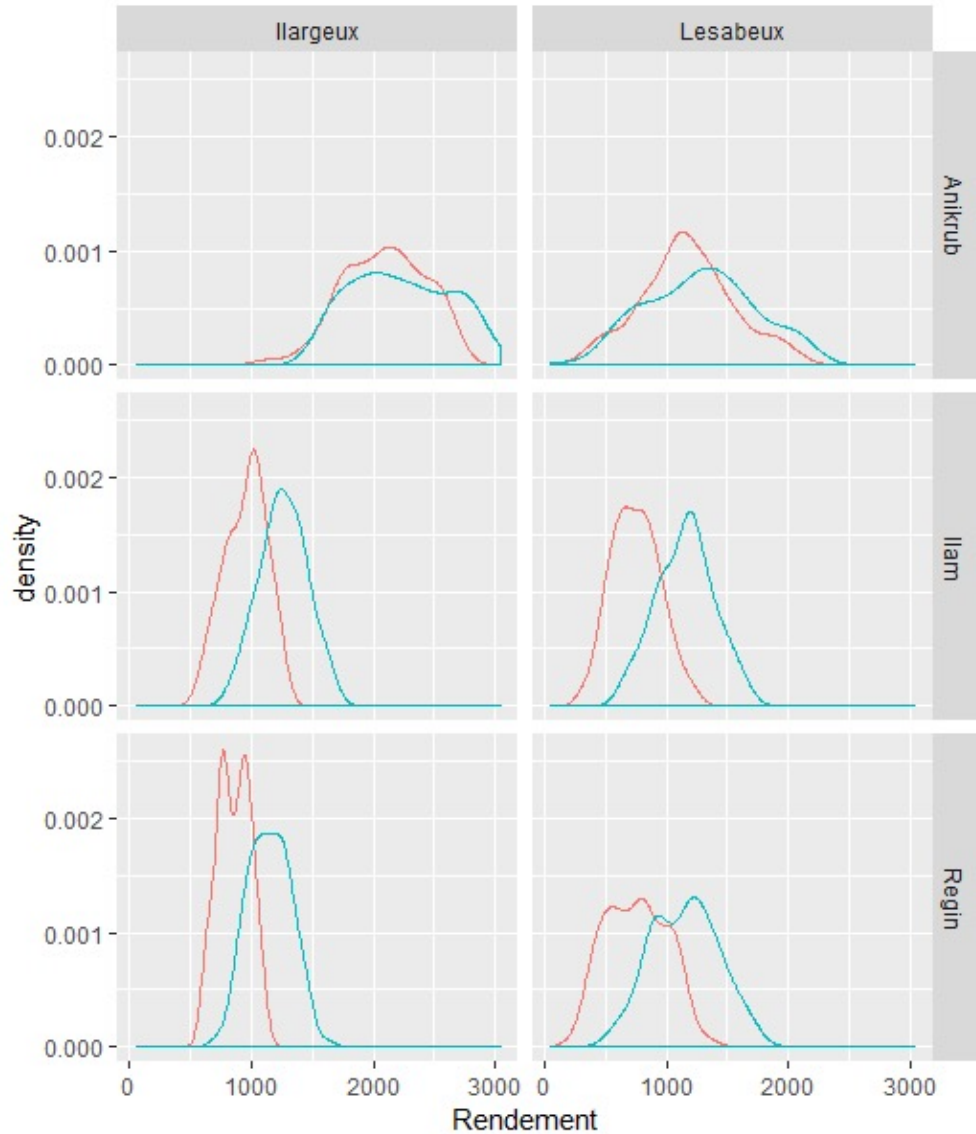
## ■ *Experimental games:*

- 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

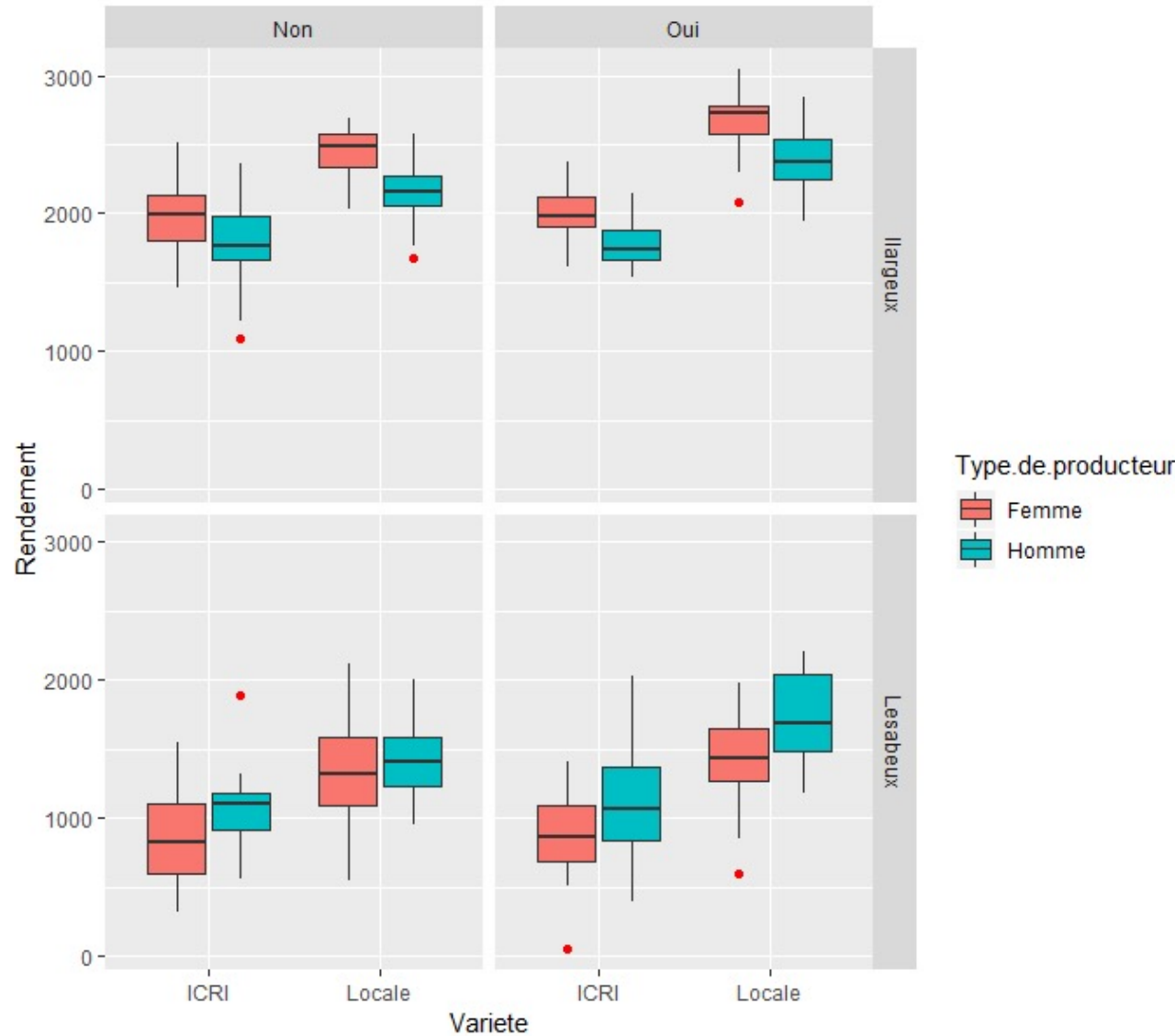
## ■ *Survey game:*

- 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



ZAI  
Non  
Oui



Type de producteur  
Femme  
Homme

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

## Question

The photograph below shows a crowd of soccer supporters at a match.

The mean weight of the supporters is 80 kg; guess their approximate standard deviation.



Approx standard deviation is  kg

Check

Tell me

Another question

## Message

Guess the standard deviation of the weights of the supporters then type it into the text-edit box above.

# What is standard deviation?

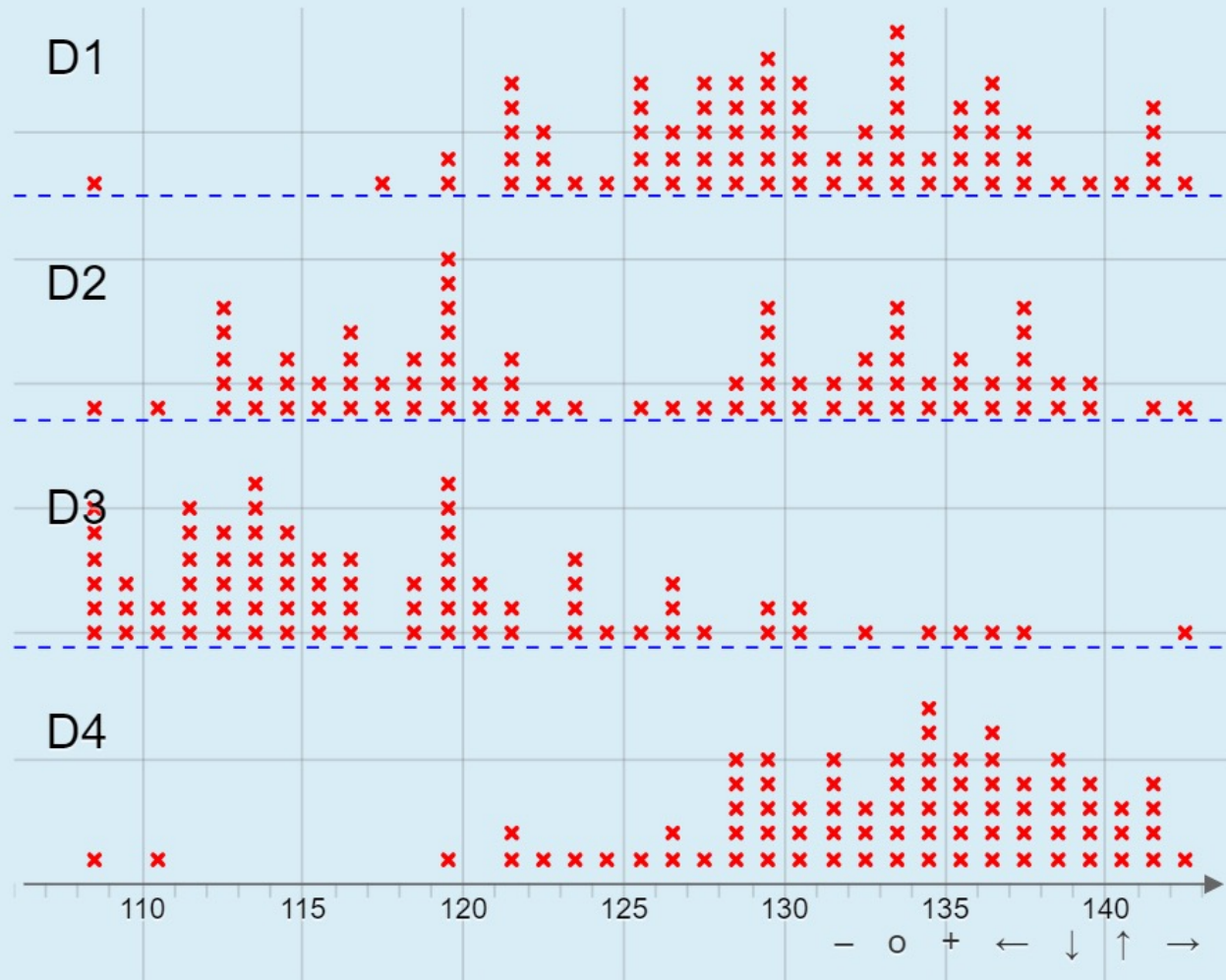
Source: CAST Computer-Assisted Statistics Textbooks, Doug Stirling  
<http://cast.massey.ac.nz/>

# Random datasets

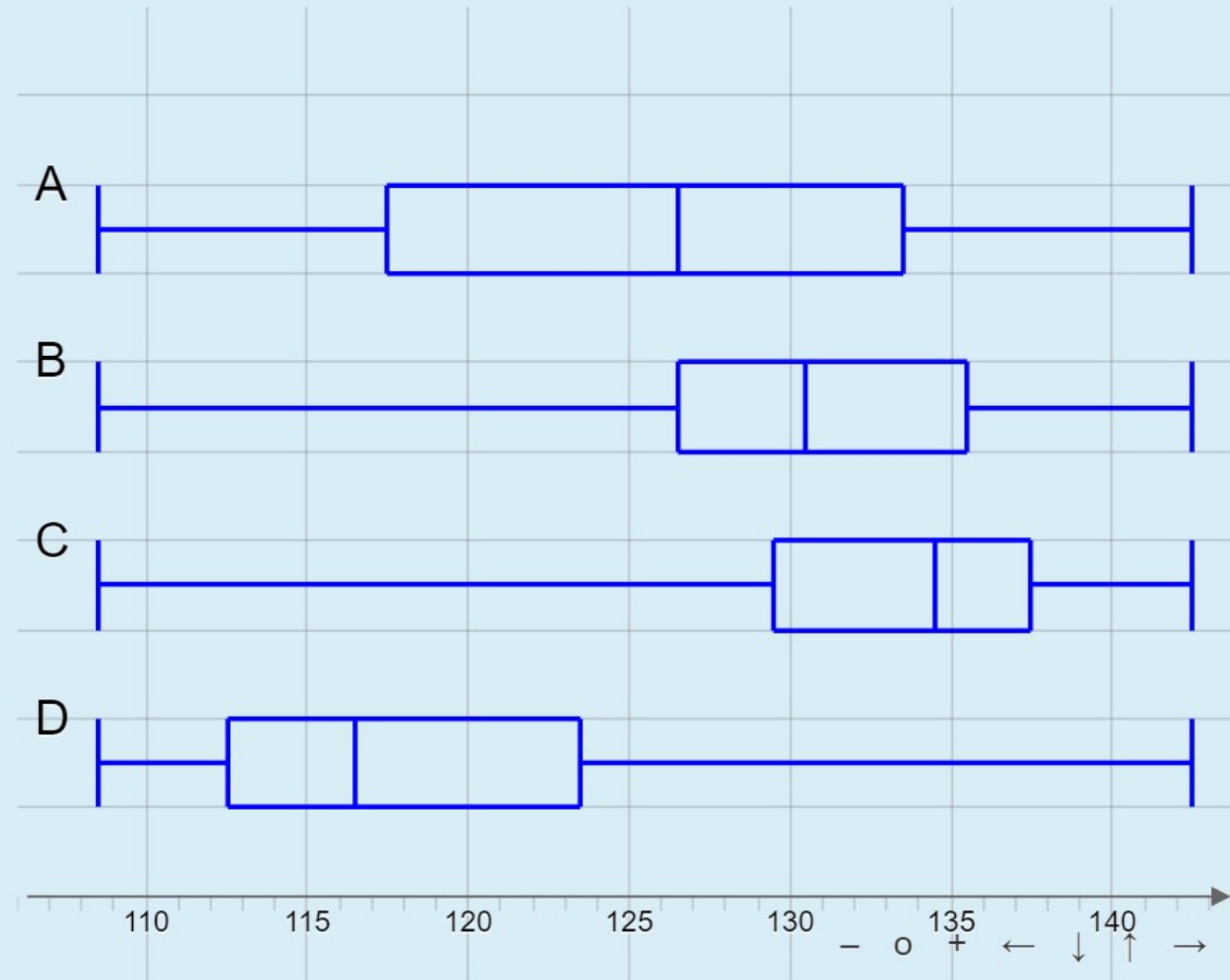
Below are stacked Dot Plots and Box Plots for four different sets of data.

Which box plot matches with each dot plot?

Dot plots:



Box Plots:



# 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.

# Student Feedback

## 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)

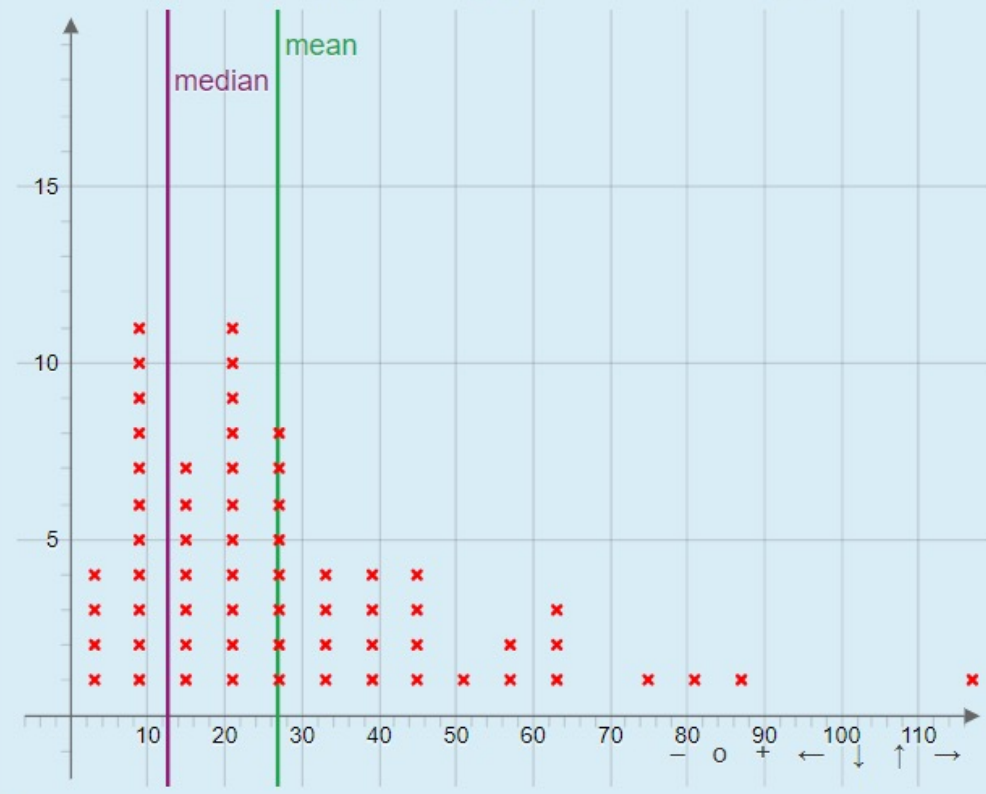
- System for automated assessment of mathematics
- Functionality of a computer algebra system (Maxima)
- Beyond multiple choice – mathematical expressions evaluated by computer algebra
  
- Works in Moodle (course management system)
- Used in hundreds of sites in multiple languages
- Free and open source ([https://github.com/math/moodle-qtype\\_stack](https://github.com/math/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

# Estimating mean & median

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.



# Detailed and specific feedback

**✘** 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.

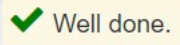
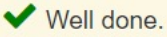
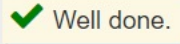
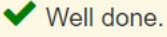
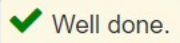
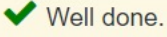
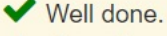
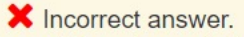
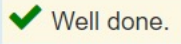
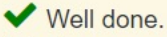
# Specific Feedback & Intelligent marking

Tidy question | Question tests & deployed versions

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

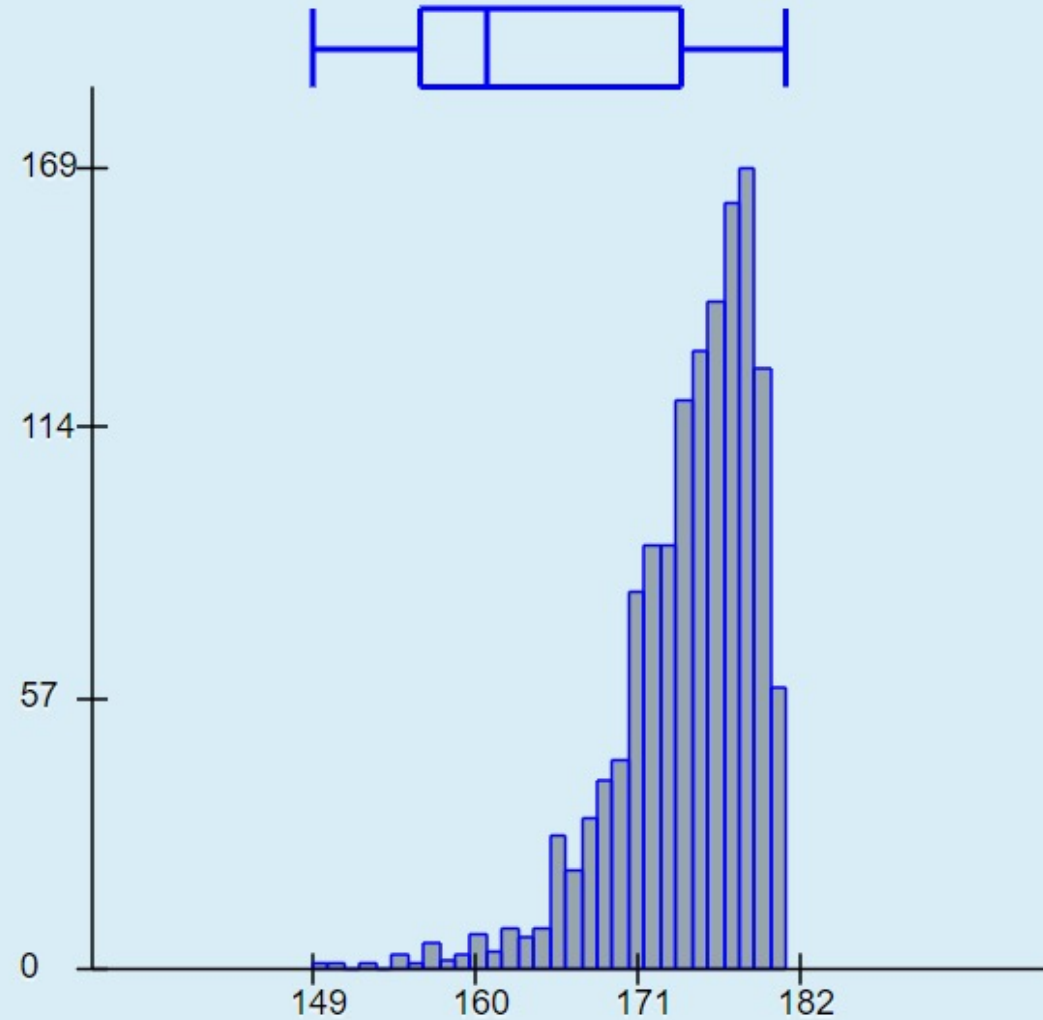
Complete the ANOVA table.

Give the values to 1 decimal place.

Term	Df	Sum Sq	Mean Sq	F Value
Light	1	9.5	9.5 	0.8  Correct F value based on your mean sum of squares and residual sum of squares.
Heat	1	46.1	46.1 	4.0  Correct F value based on your mean sum of squares and residual sum of squares.
Variety	1 	15.8	15.8  Correct mean square value based on your sum of squares and Df.	1.4  Correct F value based on your mean sum of squares and residual sum of squares.
Residuals	10  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.	114.7 	11.5  Correct residual mean square value based on your residual sum of squares and Df.	
Total	23	186.1		

# Interactive Graphics

Move the quartiles 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

(Mathematics questions)

Establish mathematical properties – infinite correct answers

Find a vector  $\mathbf{v}_4$  so that together the four vectors

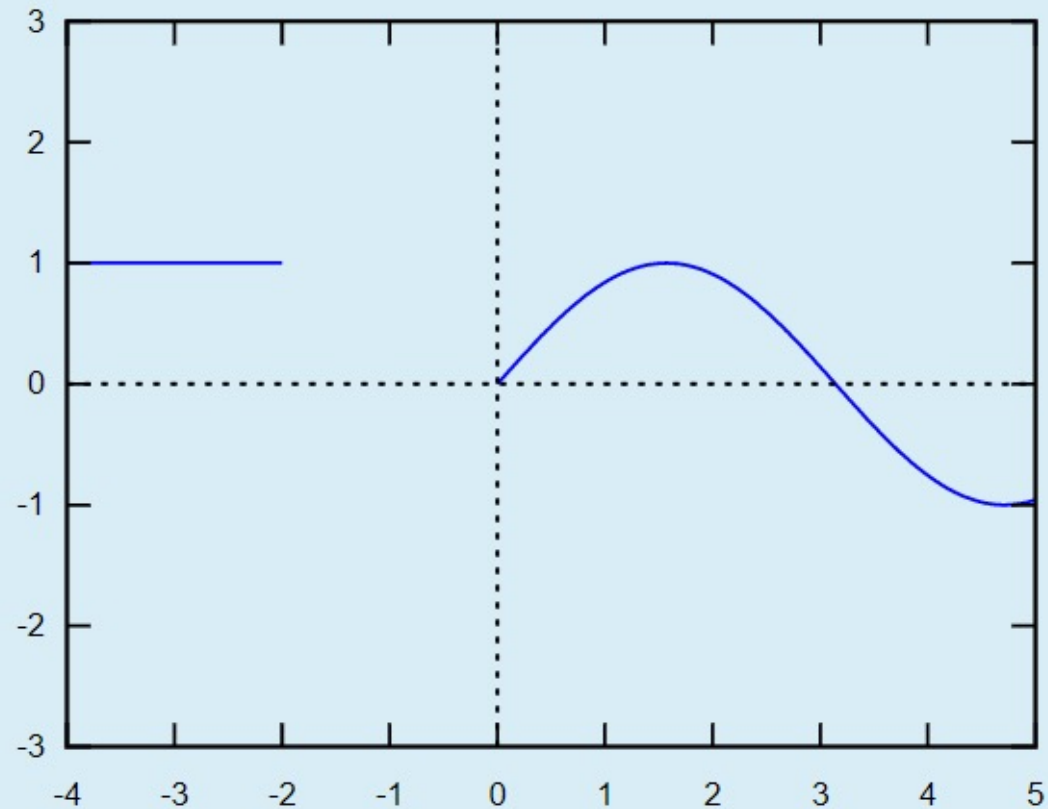
$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \mathbf{v}_4 \right\}$$

do not form a basis of  $\mathbb{R}^4$ .

<input type="text"/>
<input type="text"/>
<input type="text"/>
<input type="text"/>

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.

$$p(x) = \frac{x^2(x+3)}{4}$$



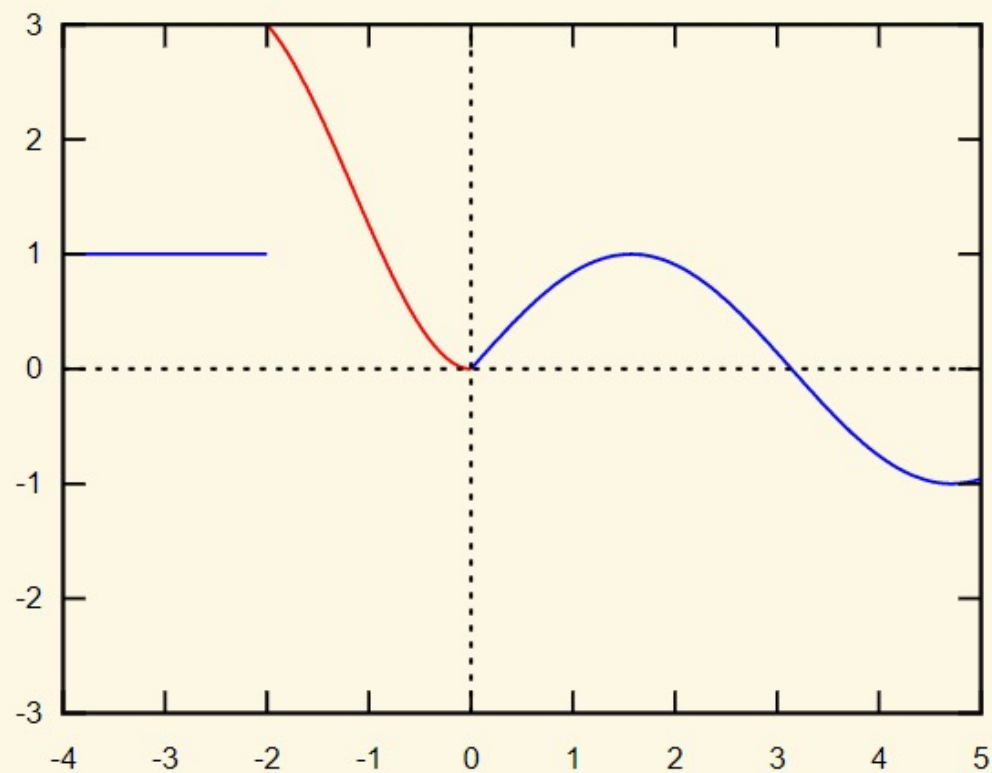
**I** Your answer is partially correct.

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

# Quiz format

	Mastery Quizzes	Test Quizzes
Attempts allowed	Infinite	1
Deadline	End of Semester	End of each week
Feedback	Immediate	End of each week
Restriction	None	>70% on Mastery Quiz
Percentage of Continuous Assessment	50%	50%

## 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

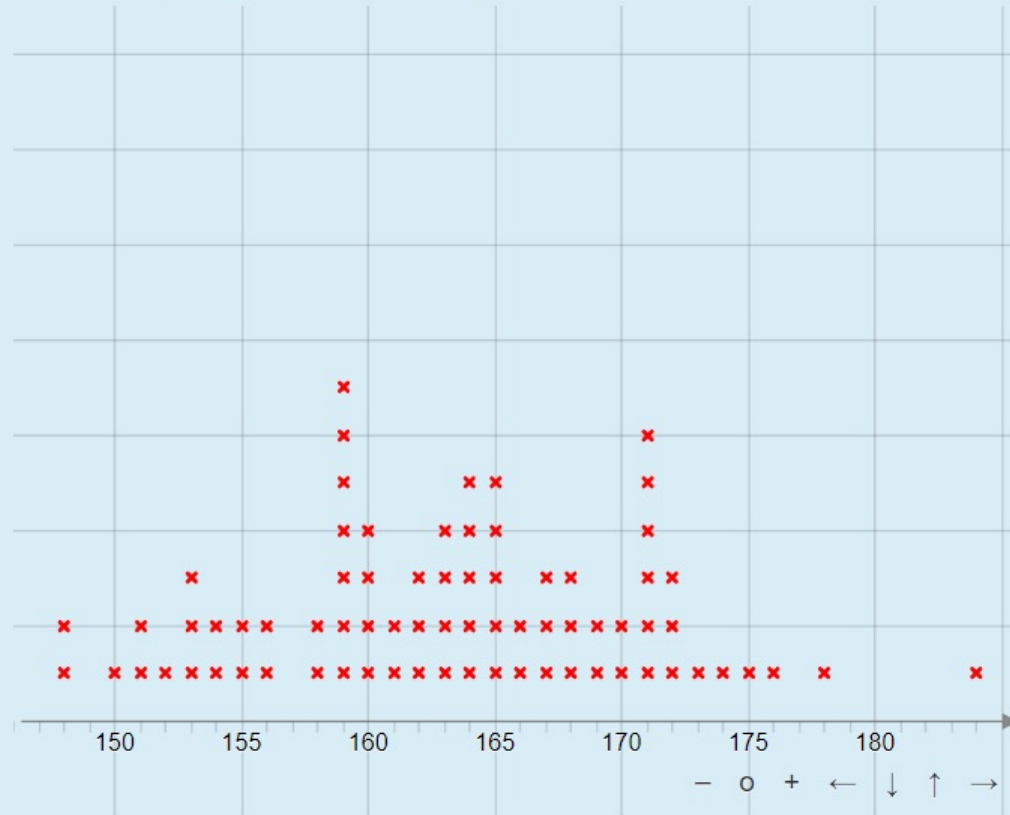
# Challenging aspects for students

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

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

- Mead, R., & Stern, R. D. (1973). The Use of a Computer in the Teaching of Statistics. *Journal of the Royal Statistical Society. Series A (General)*.
- Parsons, D., Stern, R., & Stern, D. (2019) A Problem-solving Course In Statistics For Mathematical Science Students, *Proceedings of the Satellite conference of the International Association for Statistical Education (IASE), Decision Making Based on Data*, Kuala Lumpur, Malaysia.
- Pelkola, T., Rasila, A., & Sangwin, C. (2017). Blended Mastery Learning in Mathematics. arXiv preprint arXiv:1712.07848.
- Sangwin, C. (2013). *Computer aided assessment of mathematics*. OUP Oxford.
- Sangwin, C. (2015). Computer aided assessment of mathematics using STACK. In *Selected regular lectures from the 12th international congress on mathematical education (pp. 695-713)*. Springer, Cham.
- Stern, D., Ongati, N. O., Agure, J. O., & Ogange, B. (2010). Incremental modernization of statistics teaching and curriculum at Maseno University, Kenya. *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*
- Stern, D. (2017). Seeding the African Data Initiative. *Proceedings of the IASE Satellite Conference “Teaching Statistics in a Data Rich World”*. Rabat, Morocco.
- Stern, D., Latham, S., & Stern, R. (2009). Statistical Games to support problem-based learning. *IBS 2009 SUSAN Conference proceeding*.
- Stern, D., Stern, R., Parsons, D., Musyoka, J., Torgbor, F., & Mbasu, Z. (2020). Envisioning Change in the Statistics-Education Climate. *Statistics Education Research Journal*, 19(1), 206-225
- Stern, D., Stirling, D., Dale, I., & Stern, R. (2010). The use of computer-based tests to consolidate statistical concepts in Kenya. *In ICOTS 8 Conference proceedings*.
- Stirling, D. (2010, July). Improving lectures with CAST applets. *In Data and context in statistics education: Towards an evidence-based society.-Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8)*.



## 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.