

# A NEW, INTERACTIVE AND ENGAGING WAY TO TUTOR TRAINEE SURGEONS IN STATISTICS

Riinu Ots (R.Ots@ed.ac.uk), Paula J.W. Smith, Catherine A. Shaw, Uzma Tufail-Hanif, Ewen M. Harrison, Stephen J. Wigmore, O. James Garden  
University of Edinburgh, Clinical Surgery, Royal Infirmary of Edinburgh, Edinburgh, United Kingdom

## OUR SETTING

*The Edinburgh Surgical Sciences Qualification (ESSQ)*

Since 2007, the University of Edinburgh and the Royal College of Surgeons (RCS) of Edinburgh have launched seven successful online Masters Programmes for surgeons in training: two MSc Programmes (content aligned with the intercollegiate Membership of RCS examination), and five ChM Programmes preparing for Fellowship of the RCS examinations.

The ShinyStats is a series of web-based interactive applications created using the R statistical programming language. Students are given already visualised sample data to interact with (e.g. they can change the number of observations). For each configuration, the application displays the p-value (an indicator of statistical significance).

## TAKE HOME MESSAGE

We are conscious that statistics as a subject is not popular amongst many surgical trainees, with many experiencing “statistics anxiety”. Thus, we adopted a constructivist approach in which hands-on activities and visualised interactive datasets enabled students to construct new knowledge and become more independent learners and problem solvers.

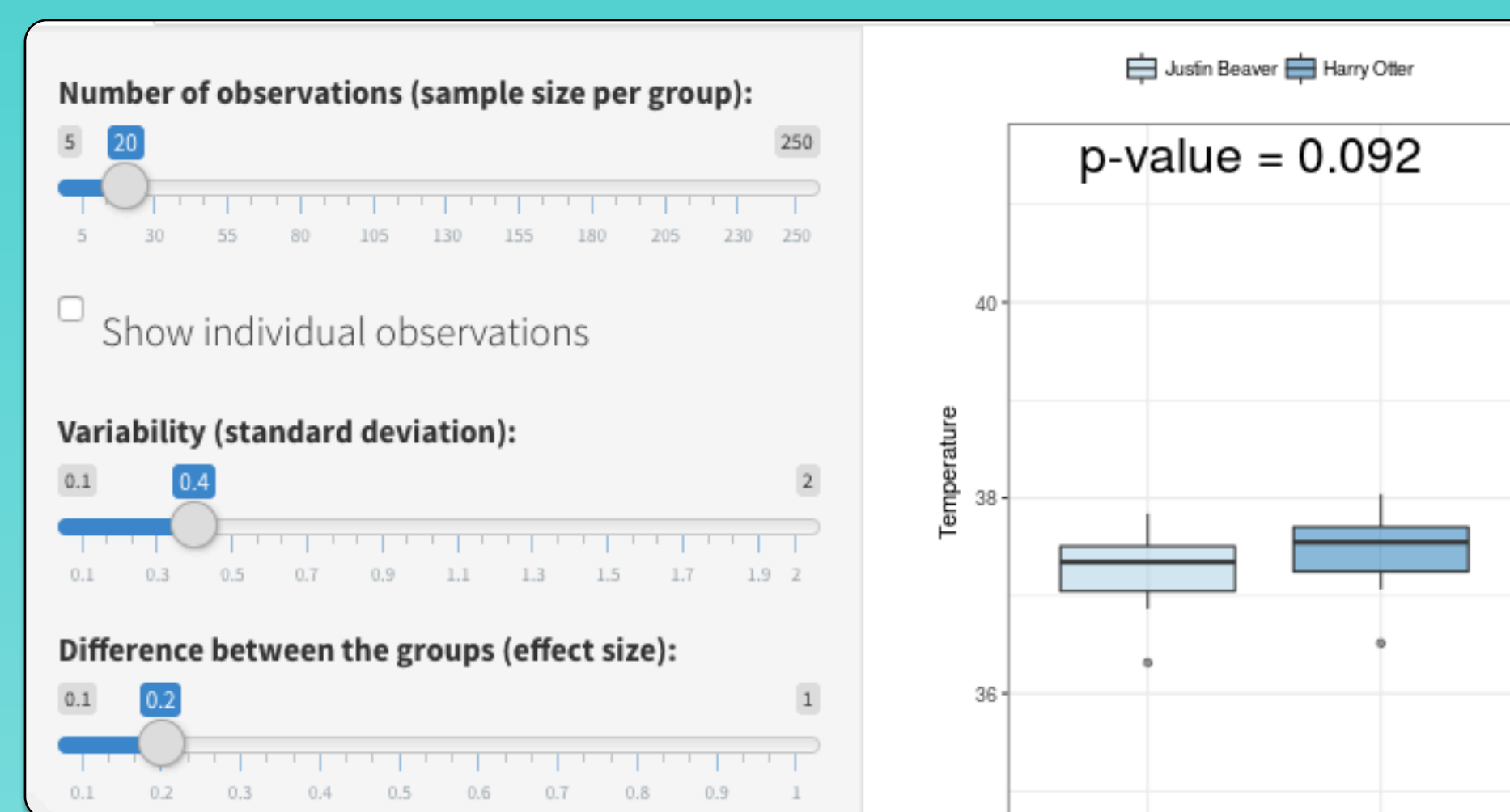


Figure 1: A screen grab of the t-test application.

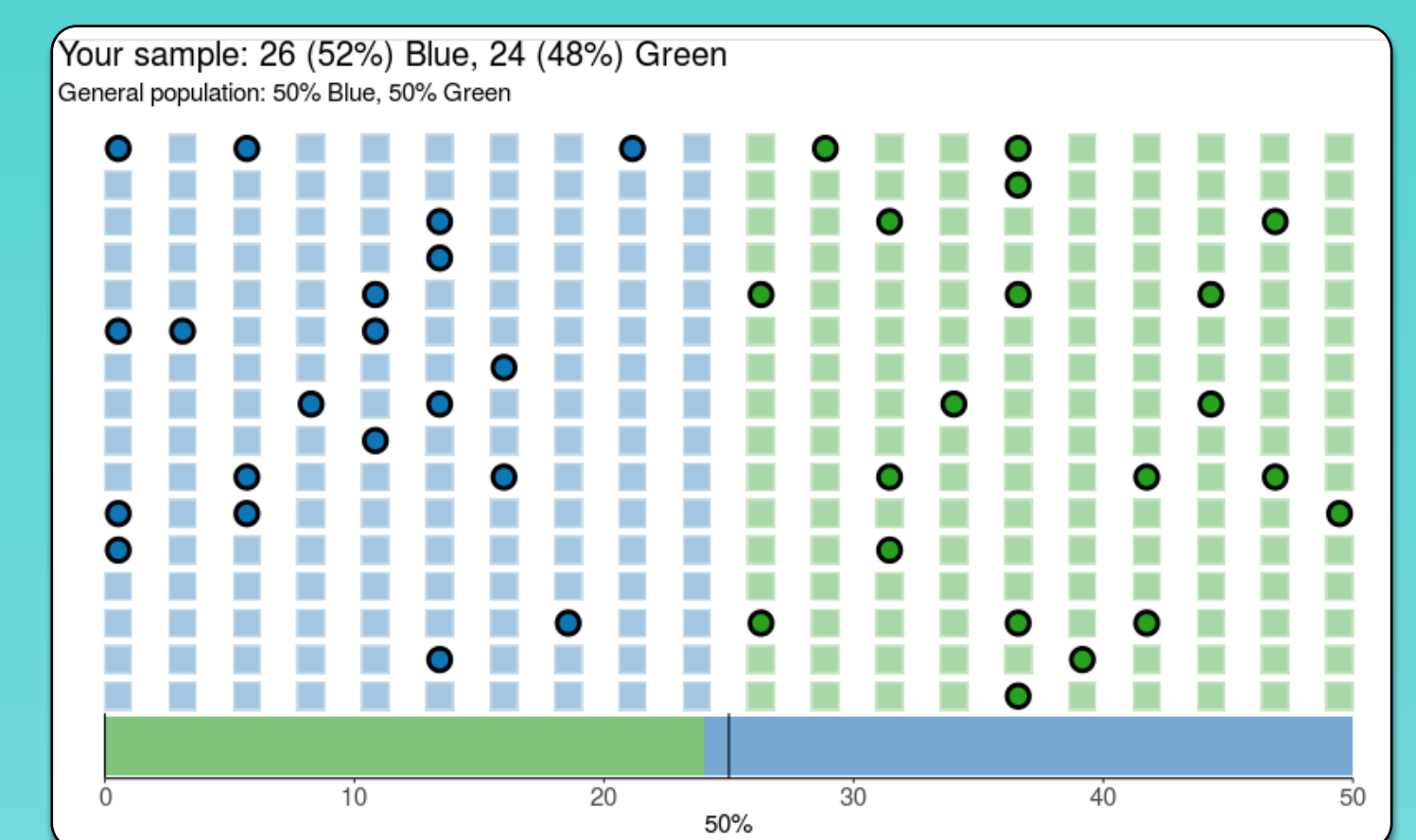


Figure 2: A screen grab of the Chi-squared application.

## EXAMPLE: POWER

With our t-test ShinyStat (Figure 1), the different values of each input combination are fixed. This means that moving the sliders to change the parameters will always show the exact same results when coming back to the original settings. By interacting with the t-test application you should really get a sense of how the sample size, standard deviation, and effect size relate to the p-value in the absence of any uncertainty that randomness might introduce.

Our chi-squared ShinyStat (Figure 2), however, does include randomness. This means that the distribution of the two groups within your sample changes every time you press the Resample button. Now, combining both of these experiences, we are ready to look at Statistical Power (the probability of finding an effect when one exists).

- First, go to the link (Figure 3), look at the power value (given in the text next to the Re-sample button), then the p-value. Then Re-sample a few times without changing the input parameters. Comment here whether the proportion of significant ( $p < 0.05$ ) vs non-significant differences is in line with the power value.
- Second, how does the 95% confidence interval relate to the p-value presented at the top? Can you guess whether the p-value is less than 0.05 by just looking at the confidence interval?

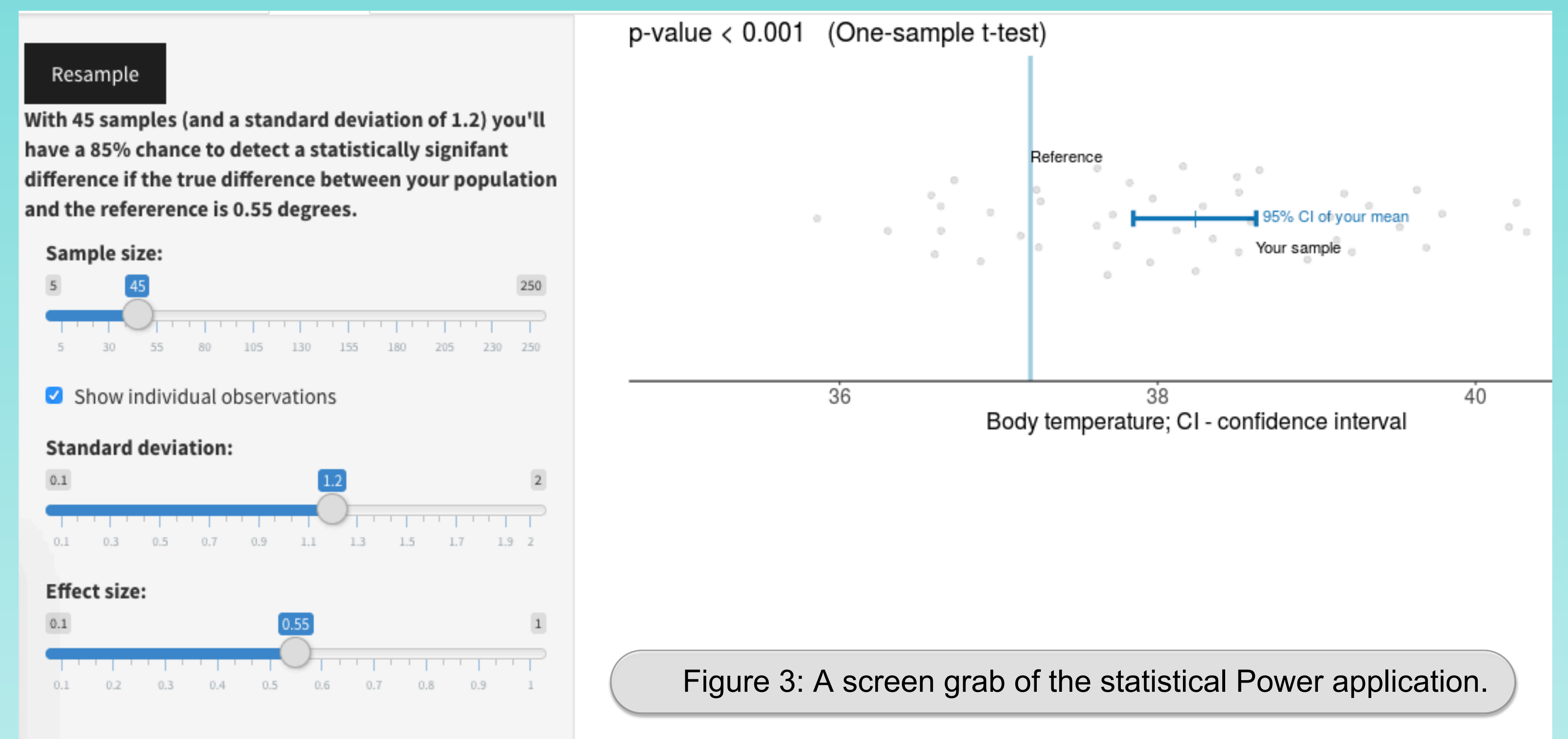


Figure 3: A screen grab of the statistical Power application.

## DISCUSSION BOARD RESPONSES:



RE: T-TEST

By moving three-slides, the number of individual observations increases, therefore the sample size increases. Therefore the probability that the result is due to chance reduces, resulting in a smaller p value and the more significant is the result.



RE: BoxPLOTS

Playing around with it some more, the the single point plotted corresponds to the outliers that do not fall on the the whiskers of the box plot, outside of 1.5X the width of the IQR. On the app, with the show individual observations, it corresponds to either the highest or lowest recorded values outside of the whiskers range. The more I increase the variability, effect size and number of observations, the more dots appear, again in line with the outliers.  
Like [other student's name], stats is also not my strong suit.



RE: SAMPLE SIZE

1. minimum number of observations needed to achieve significance with standard deviation of 0.4 and effect size of 0.2 is 30  
2. smallest difference between groups needed to give a statistically significant value with sample size of 30 and standard deviation of 0.5



RE: MULTIPLE TESTING

Sample - 66% blue and 34% green  
p= 0.024 so significant  
It took 36 attempts at resampling!  
Is it that if you sample a population enough time - you can get a chi-squared test to be "statistically significant" - so in reality the sample isn't genuinely different to the population; but we sampled it enough times that it appears to be so?



RE: MULTIPLE TESTING

p of 0.05 is the same as saying that 5% of the time, the result would be by chance (or if one took a sample 100 times, 5 times it would be coincidence).  
Therefore with the example, one should click 20 times (5/100) in order to get a chance result.



RE: MULTIPLE TESTING

Agreed, a lot simpler than what I was trying to find a calculation for. Thanks [other student's name] and Riinu