"Why do we teach so much Statistical Hypothesis Inference Testing"

Paul Hewson
paul.hewson@plymouth.ac.uk

May 2012

Teaching Statistical Hypothesis Testing

- How many threshold concepts (e.g. Dunne *et al.* 2003) do we need to be comfortable with in order to really understand hypothesis testing?
- Sure, we can "beat" students into passing exams, but what do we achieve "if the only tool you have is a hammer, every problem looks like a nail"?

Randomisation tests

- Arguably, these pre-date more “mathematical” methods
- But very few texts describe them
- They are examined on very few syllabi
- But list out the threshold concepts needed in order to understand them

An experiment

- Link between Popper's philosophy (falsifiability) does make H_0 rejection conceptually more straightforward with designed experiments
- Plenty of simple experiments lend themselves to within lecture settings (we looked at cola and stress levels)
- Write null, perform experiment, collect data, measure observed test statistic (so far so standard)

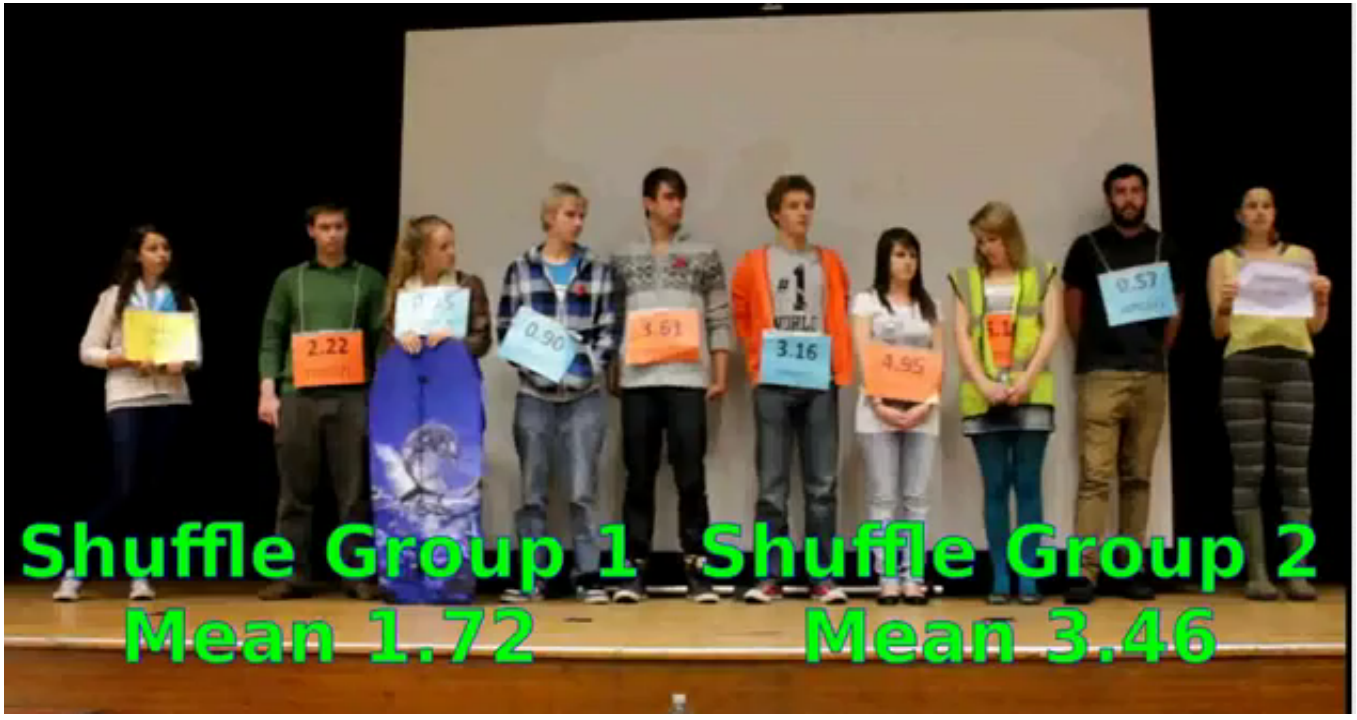
Primary data collection

But if H_0 were true

(kind of)

- These data are more a feature of the participants than the treatments
- We could have randomly allocated participants to the other treatments
- We would have got those measurements attributed to other treatments

So if H_0 were true we could shuffle the data (and repeat)



What we get

- An observed test statistic
- A number of random shuffles (each of which is plausible under the null) - and an associated discussion about what that means
- A chance to discuss what we regard as evidence to reject the null
- First attempt to capture this is here: www.youtube.com/watch?v=ESP0huKsKD0

OK, but this is all really obvious and lots of people do this

- Yes I know . . . everyone who sees this poster knows this already, does this already (possibly taught / inspired me to do it)
- *So why isn't it standard in texts and on syllabus sheets?*
- Why do we get the consulting questions we get