



Viewpoint

Beware multiple hypothesis testing—the curse of type I errors

Journal of the Operational Research Society (2006) 57, 612.
doi:10.1057/palgrave.jors.2602127

I have been prompted to mount my OR educator's soapbox, if only briefly, by Dr Koh's reply (Koh, 2004a) to my note (Worthington, 2004) on her paper (Koh, 2004b).

First of all, I would like to reassure readers that, contrary to the implication of Dr Koh's reply (Koh, 2004a), the laws of statistics (in particular, the law implied by the title of this viewpoint) do not cease to hold because the SIMAN random number generator (good as it may be) has been used to generate a set of experimental results.

I know that I am not the only OR educator who warns their students of the ease and dangers of misusing advanced software packages, be they for simulation, statistical analysis, linear programming, etc. Koh's original paper (2004b), its progress into *JORS*, and her response to my previous letter are all evidence that even academics with their reflective inclination and access to expert advice can be lulled into a false sense of security when using user-friendly packages.

As my co-workers have been quick to point out to me, the statistical issue identified in my previous note was 'not rocket science'. It goes under the very unexciting name of 'type I errors' and will be taught to my second year management students with renewed vigour when we look at hypothesis testing later this term. When testing a hypothesis at a (say) 5% significance level,

$$\text{Probability (type I error) = probability (rejecting the } H_0 \text{ when } H_0 \text{ is true) = 5\%}$$

When an analyst or researcher reports the results of a single-hypothesis test as significant at the 5% level, this is usually simply interpreted as a small risk (ie a 5% risk) of being wrong. However, when many hypotheses are tested (Koh's paper included 92), the 5% significance level becomes a virtual guarantee that the analyst will draw a number (ie about 4 or 5 in this case) of incorrect conclusions. This issue needs to be recognized and the results need to be interpreted accordingly and/or backed up with further experiments as explained in my previous note (Worthington, 2004).

And so onto my soapbox:

- Fellow OR educators should not lose sight of the need to recommend that advanced packages are used at least with care, and where possible with some friendly expert advice.
- Fellow researchers and analysts should continue to make interesting and valuable use of simulation modelling, as in Koh (2004b), but beware of the challenge of analysing the results soundly.

References

- Koh SCL (2004a). Viewpoint: comment on Dave Worthington's note on Koh SCL. MRP-controlled batch-manufacturing environment under uncertainty. *J Opl Res Soc* 55: 1235.
- Koh SCL (2004b). MRP-controlled batch-manufacturing environment under uncertainty. *J Opl Res Soc* 55: 219–232.
- Worthington D (2004). Viewpoint: a note on SCL Koh's paper. MRP-controlled batch-manufacturing environment under uncertainty. *J Opl Res Soc* 55: 1235.

Lancaster University, UK

D Worthington