Applet #4 Activities: Instructor Guide

Learners studying introductory statistics usually find concepts relating to hypothesis tests difficult to understand. One reason for this is that how a test statistic varies over multiple samples is not easy to convey in a textbook or traditional lecture. A variety of methods have been proposed for aiding learning about hypothesis tests, one being on–line simulation–based tools, or *applets*, that permit visualisation of multiple samples from a distribution and how test statistics vary. Research suggests that how students engage with such applets can impact the learning obtained, as without structure to their interactions with a visualisation tool learners may miss important concepts.

Prof. Mike Whitlock (in the Department of Zoology, University of British Columbia) has developed a suite of applets for use in introductory courses. The fourth of these, "Chi-squared contingency analysis", aims to enhance understanding of contingency tables and hypothesis testing via simulations of tables and the Chi-squared goodness–of–fit test statistic. The accompanying activity aims to assist the learner in their engagement with the applet, helping to focus attention on key learning goals. The activity may be used as an in–class activity (say as part of a lab–based tutorial) or outside class as a homework assignment. There follows an instructor guide to the activity.

The activity targets two sets of learning aims: (i) appreciation of exploring relationships between categorical variables via contingency tables and (ii) understanding aspects of hypothesis testing, including P-values, type I and II errors, and power. Instructors wishing just to target aim (i) may use only parts 1, 2, 3, 8, and 9 of the activity.

Part 1 probes if the learner understands the idea of independence between the two variables – sometimes termed *homogeneity* in contexts where the marginal distribution of one categorical variable does not change across different subpopulations. Part 2 clarifies the definition of the total sample size n and asks the learner to predict what a possible table may look like given the default settings of n and the parameter values. In 3 a table is created from simulated data, which may differ from what the learner anticipated either due to a misunderstanding of independence or a failure to appreciate sampling variability.

The χ^2 test statistic is informally introduced in part 4. Definitions of the null hypothesis and P-value are given. The learner's attention is focused on whether the table they sampled resulted in the rejection of H_0. Part 5 repeats the process, drawing attention to the sampling distribution of χ^2

and the observed P-value.

Asking the learner for a prediction prior to observing a simulation is known likely to enhance learning from simulation–based tools. In part 6 the learner is asked to predict the number of false rejections in fifty simulations. It is hoped any learner who was initially confused about independence has realised that the null hypothesis holds for the default parameters. The picture should become clear once fifty simulated tables are analysed; this occurs in part 7, where type I error is introduced.

Parts 8 and 9 repeat 2 and 3 with the parameters changed so there is dependence between the two variables. Type II error is introduced in part 10, though unlike with type I error in 7 it is not possible for the learner to predict the rate of type II errors. However, the rate of false rejections can be estimated from the fifty simulations observed.

In part 11 the power of the test is defined and the parameter values changed to make the dependence stronger, that is, the two marginal distributions differ more than in part 10. The learner should appreciate this change increases the power of the χ^2 test. Parts 12 and 13 illustrate that the power of the test will decrease as n decreases, while in 14 the parameters are re-set to maximise the dependence, which the learner may anticipate will increase the power; either way, part 15 illustrates via simulation. Finally in 16, the learner should now understand that increasing the sample size back to 120 will increase the power again.

Remark 1 *Very rarely in multiple simulations it may become apparent that the number of χ^2 values in the critical region and which appear red in the graphic does not tally with the count in the yellow box. This is due to the axis not rescaling in the event of an extreme value of χ^2 occurring that is to the right of the maximum value on the horizontal axis. Even when such an issue occurs it can be nearly impossible to detect. Rest assured that the counts displayed are accurate.*

Remark 2 *Selecting n odd always results in a sample of $n/2$ from the unvaccinated subpopulation and $n/2+1$ from the vaccinated. This could perhaps distract learners so it is advised, initially at least, to keep n even. It should also be noted that stratification is implicit in the sampling scheme applied for each simulation.*