Biostats VHM 801 Course
Fall 2019, Atlantic Veterinary College, PEI
Henrik Stryhn

Supplementary exercises

The purpose of these exercises is to allow a student seeking challenges beyond the standard curriculum of VHM 801 to explore some questions of both theoretical and practical relevance.

Problem 1. Calculation of probability by simulation.

The question came up, "How likely is it for a valid column blocking scheme (e.g. below) in a 4×4 layout (with 4 treatments) to occur by chance alone?" (i.e., in a completely randomized design).

C	A	B	D
A	B	D	C
D	D	A	B
B	C	C	A

While the probability can be computed exactly using combinatorial arguments, it is useful to explore how it can be estimated (to any desired accuracy) by simulation. In your statistical software of choice, simulate a random assignment of the 16 units in the square, and determine a set of rules to determine whether the random assignment achieved meets the condition. Then repeat the process a large number of times, and compute the proportion of assignments meeting the condition. The approach would often be referred as "Monte-Carlo estimation".

Problem 2. Proportion of "suspected outliers" in different distributions.

How many observations the rule for "suspected outliers" based on the interquartile range will flag depends on the distributional shape. It is useful to have an intuitive understanding of this relationship. The task is therefore to compute the proportion beyond the cut-off of the rule for a number of theoretical distributions of different shapes.

1) Start by a standard normal distribution $\mathrm{N}(0, 1)$; the answer can be checked with the value given in Lecture 1.
2) To represent symmetrical distributions with less tails than the normal distribution, compute for the uniform distribution and the triangular distribution; both of these distributions have smaller kurtosis than $\mathrm{N}(0, 1)$.
3) To represent symmetrical distribution with heavier tails than the normal distribution, compute for t-distributions with varying degrees of freedom; these distributions have larger kurtosis than $\mathrm{N}(0, 1)$.
4) To represent right-skewed distributions, compute for log-normal distributions with varying standard deviation σ; the (positive) skewness increases with values of σ.