STATISTICAL METHODS II
ASSIGNMENT 06
SOLUTIONS

Problem 06.2 [[5]]

The biological half-life of a drug ...

Solution: Oops. The expected counts are off, but let us compare the expected and the observed anyway, which is what we should have done. This is usually done using a chi-squared test, but you do need to ensure that all expected cell counts are greater than 6, or else the assumptions of the chi-squared will not be met.

The null hypothesis is that there is no difference between the two distributions (expected and observed). The alternative hypothesis is that there *is* a difference between the two distributions. This test uses all nine pieces of information, rather than just some summary statistic (like the mean or variance). Pearson's Chi-squared test indicates that there is no appreciable difference between the two distributions, thus we must conclude that there is no significant evidence that aspirin helps to prolong the effects of caffeine in the body ($X^2 = 8.12, df = 8, p = 0.4223$). Figure 1 compares the concentrations of caffeine in the blood with, and without, aspirin. Note that the concentration line without aspirin is always above the line with aspirin. This suggests that aspirin may actually eliminate caffeine from the blood faster than without. The statistical evidence, however, only tells us that there is no appreciable difference. ◇

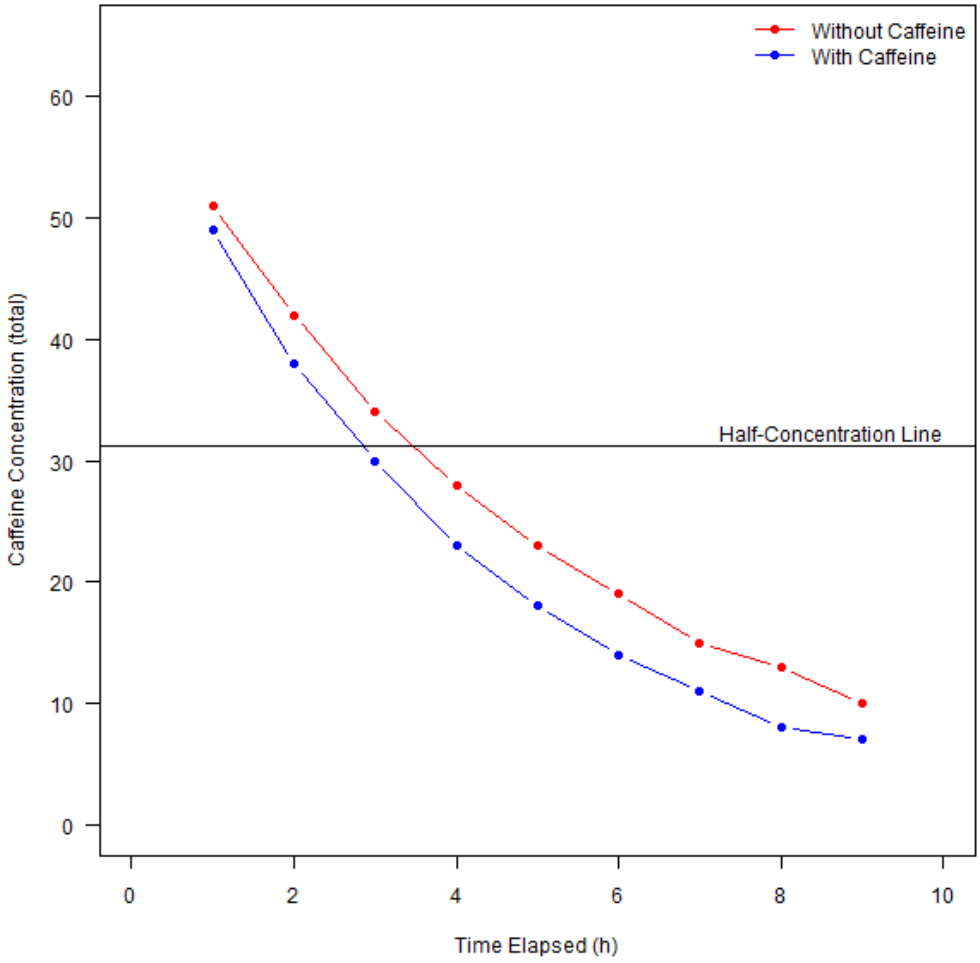

Figure 1. *Plot comparing the concentrations of caffeine in the blood.*

Problem 06.3 [[5]]

In January 2010, there was a presidential election in the island state of Sri Lanka ...

Solution: There are two ways of showing this, both deal with comparing means. If your boxplot contained only two boxes (Figure 2, top), one for provinces voting for Rajapaksa and one for those voting for Fonseka, then you would use a simple comparison of two means test (parametric or non-paramteric, depending on the assumptions). If your boxplot contained nine boxes (Figure 2, bottom), one for each province in Sri Lanka, then you would have to perform an analysis of variance-type test and then determine which provinces were not like the others (based on multiple comparisons).

The data, in either case, fails both the Normality test and the equal-variance tests.[1] Thus, we will have to use non-parametric tests (the Mann-Whitney test for comparing two means, or the Kruskal-Wallis test for comparing multiple means).

In either case, the data show that those provinces that voted for Rajapaksa had a significantly lower proportion of rejected votes than did the other provinces ($W = 4171, p \ll 0.0001$). Furthermore, if you looked at the nine provinces separately, you would find that the Northern Province is in a class by itself; Eastern, Uva, and Central are in a class by themselves; and the rest are grouped together.

This method does not give such a clear-cut conclusion. It is clear that the Northern Province is different from the rest, but the Eastern Province (the other one which did not vote for Rajapaksa) is not significantly different from some provinces that supported Rajapaksa. ◇

[1] Transforming the data may be appropriate, but the Box-Cox transformation was unable to make the data Normal and have equal variance, so no simple transformation will work.

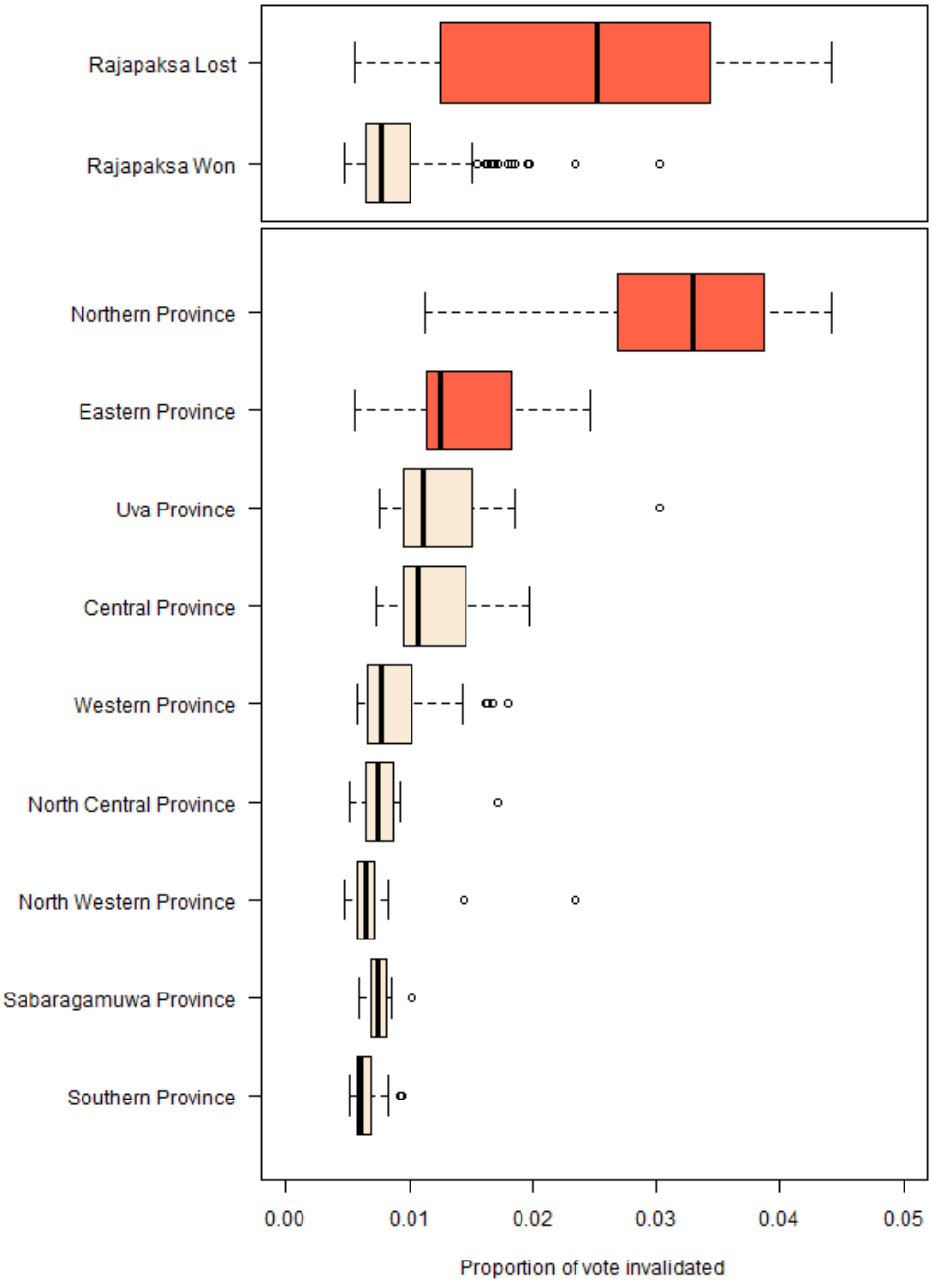

Figure 2. *Boxplot of the proportion of votes declared invalid by the province.*