MA5121: Nonparametric Statistics

Satvik Saha, `19MS154` December 5, 2023

Problem 1 Consider the following samples.

$$X : 35, 66, 58, 83, 71.$$
$$Y : 46, 56, 60, 49.$$

The two-sided Mood test of scale gives a p-value of 0.2072, and the two-sided Ansari-Bradley test gives a p-value of 0.5238. Thus, we fail to reject the null hypothesis that X and Y differ in variability at an $\alpha = 0.05$ level of significance.

Given the sparsity of the data, it is difficult to give a definitive answer, especially since the visualization below seems to indicate that X has greater variability than Y. A two-sided F-test for comparing variance (assuming that the distributions of X and Y are normal) gives a p-value of 0.1225, and a 95% confidence interval of $(0.52, 78)$ for the ratio of variances.
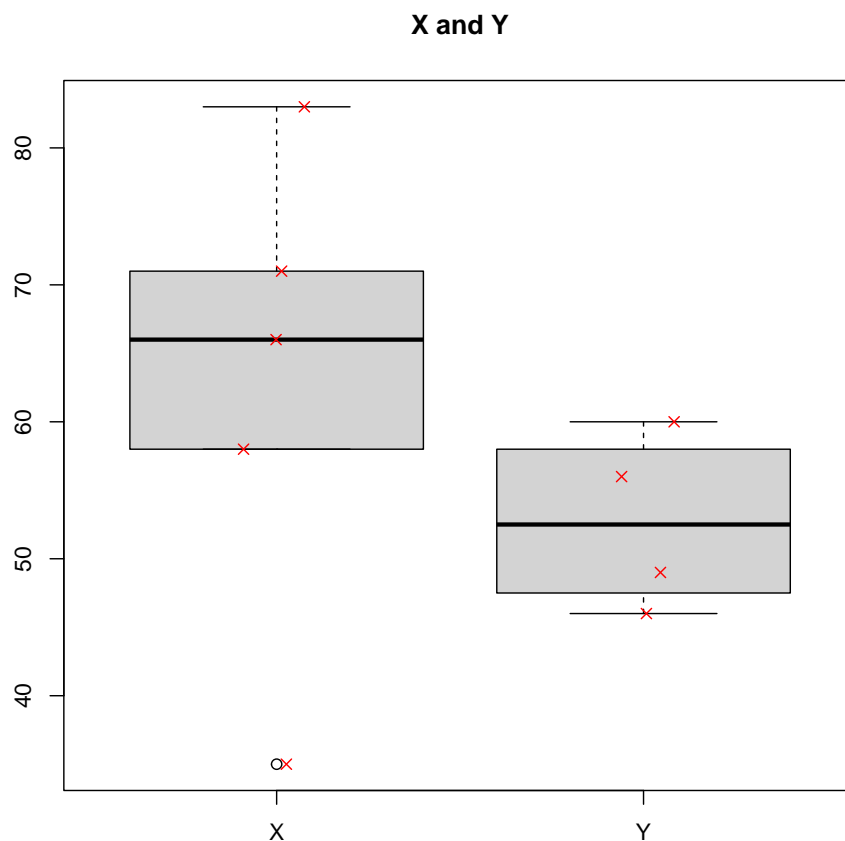

Figure 1: X and Y.

Problem 2 Consider the following distributions of number of days until death (by tuberculosis infection) for mice.

$$(\text{Control}) \; A : 5, 6, 7, 7, 8, 8, 8, 9, 12.$$
$$(\text{Treated by drug}) \; B : 7, 8, 8, 8, 9, 9, 12, 13, 14, 17.$$

We set up the null hypothesis that the distributions of A and B are identical, versus the alternative hypothesis that $A \overset{st}{<} B$.

The Wilcoxon rank-sum test gives a p-value of around 0.02, using which we reject the null hypothesis at a 0.05 level of significance. Thus, we infer that the drug treatment has the positive effect of prolonging the number of days till death by tuberculosis infection for mice.

We employ the one-sided Wilcoxon rank-sum test since we wish to determine whether the distributions of number of days (for the groups A and B) are identical or not. We also discard the possibility that $A \overset{st}{>} B$. However, it is difficult to justify the assumption that the underlying distributions are continuous since our data is quite discrete, and thus the large number of ties may make the rank-sum test unreliable.

The t-test for location, assuming normality (which can be justified to some extent via the Shapiro-Wilk test; the t-test is also quite robust to non-normality), also yields a p-value of around 0.022, with which we may again reject the null hypothesis.
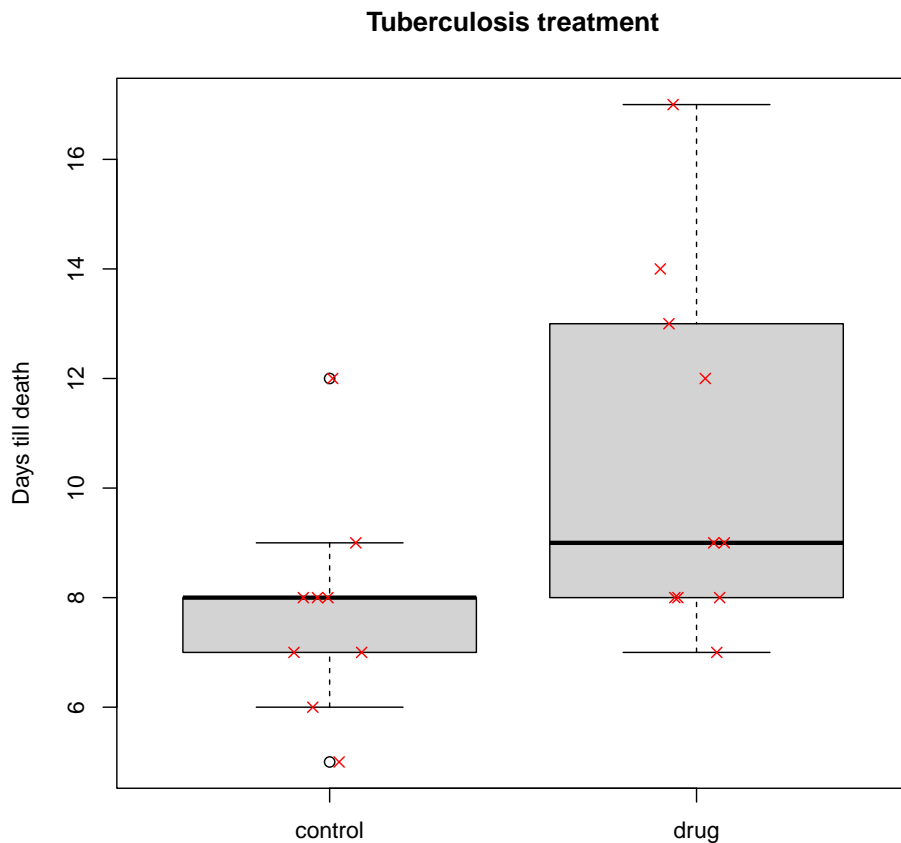

Figure 2: Days till death by tuberculosis for mice.