The Rank-Sum Test

A survey in a statistics class at Penn State University contained the question "What is the fastest you have ever driven a car? _____mph." The responses of 7 females and 7 males are provided below.

Females:	80	85	30	70	105	90	95
Males:	110	110	90	140	100	150	120

If we pretend that these 7 females and 7 males are like random samples from the populations of university women and men, respectively, we might use this data to test for a difference between the distribution of speeds that would be reported by all university women and the distribution of speeds that would be reported by all university men.

1. Are there any reasons why a two-sample t-test might not be appropriate for this data?

2. The *rank-sum test* (also known as the *Wilcoxon test* or the *Mann-Whitney test*) can be used to test for a difference between two population distributions or to test for a treatment effect in the case of a randomized experiment. Unlike the two-sample t-test, the *rank-sum test* does not require roughly normal distributions. The *rank-sum test* performs nearly as well as the two-sample t-test when population distributions are normal and can perform much better than the t-test when there are extreme outliers. We will conduct the *rank-sum test* with the speed data.

   (a) Begin by listing the observed data values (observations for short) from smallest on the left to largest on the right. Use all $n_1 + n_2 = 7 + 7 = 14$ observed data values and make no distinction between the two groups at this point.

   (b) Number the observed data values from 1 (smallest) to $n_1 + n_2 = 7 + 7 = 14$ (largest).

   (c) Assign a rank for each observed data value. The rank of an observed data value is simply the number assigned to it in step (b) unless the observed data value appears more than once in the combined data set. For each duplicated data value, find the average of the numbers assigned to the data value in step (b). This average is used as the rank for each observation that shares the data value.

   (d) Label each rank with F or M depending on whether the rank belongs to a female or male observation.

   (e) Find the sum of the female ranks. Call this quantity T.

(f) If there is no difference between the distribution of speeds driven by university women and men, the ranks assigned to the female observations should be like a random sample of n_1 ranks from the group of ranks determined in step (c). One way to determine if there is significant evidence of a difference between the distributions would be to compare the observed value of T with the values obtained by computing T for all possible samples of n_1 of the $n_1 + n_2$ ranks determined in step (c). If the sample sizes are not too small, we can approximate the p-value that we would obtain using this procedure by comparing the test statistic Z defined below to a standard normal distribution.

$$Z = \frac{T - \text{Mean}(T)}{\text{SD}(T)}, \text{ where } \text{Mean}(T) = n_1 \bar{R}, \text{SD}(T) = s_R \sqrt{\frac{n_1 n_2}{n_1 + n_2}},$$

$\bar{R}$ is the mean of the ranks determined in part (c), and s_R is the standard deviation of the ranks computed in part (c). Compute Z for the speed data, given that $s_R \approx 4.174$.

(g) If the two population distributions are identical (or, in the case of a randomized experiment, there is no treatment effect), the distribution of Z is approximately standard normal. (This requires reasonable sample sizes in each group and few ties.) Use this fact and the value of Z computed above to find a p-value that summarizes the strength of evidence against the null hypothesis of no difference between the distribution of speeds driven by university women and men.

(h) Does there appear to be a difference in the distributions? Provide a conclusion for your test.