Exam 1: March 3rd

Sheet of paper, notes on front and back, handwritten

Let the observations from treatment 1 be denoted X_1, X_2,..., X_m and the observations from treatment 2 be denoted Y_1, Y_2,...,Y_n.
Let us assume that the data have no ties, so that any given observation is either strictly less than or strictly greater than any other observation.
Mann-Whitney statistic: Denoted U, is defined as
U = number of pairs (X_i, Y_j) for which X_i < Y_j

The null hypothesis is that the distributions of the X's and the Y's
A large value of U indicates that the larger observations tend to occur with treatment 2 (the Y's), and vice versa if U is small.
Table A4: Lower-tail and upper-tail values for the distribution of U under the null hypotheses are given in the Appendix.

- Upper-tail and lower-tail values are related by the equation
U_upper = m*n - U_lower

For each the 16 pairs (X_i, Y_j), U = 12 satisfy the inequality X_i < Y_j

The Mann-Whitney statistic can be shown to be equivalent to the Wilcoxon rank-sum statistic in the sense that one is a linear function of the other.
R(Y_j) = (number of Y's <= Y_j) + (number of X's <= Y_j)

Let's assume that the Y's have been arranged from smallest to largest: Y_1 < Y_2 < ... < Y_n.
Let W_2 denote the sum of the ranks of the Y's.
Since:

$$Y's \le Y_j = j$$

Then:

$$W_2 = \sum_1^n R(Y_j) = 1 + 2 + \dots + n + U$$

So:

$$W_2 = \frac{n(n+1)}{2} + U$$

For certain theoretical purposes.......

Let F_1(x) and F_2(x) denote the cdf's of the X's and the Y's respectively.
If there is a difference between treatments, suppose that the effect is to shift the distribution of one treatment an amount Δ to the right or left of the distribution of the other treatment.
F_1(x) = F_2(x-Δ)
Note: Can think of Δ as the difference between the means (if they exist) or medians of the two distributions.

Since:

$$P(X_i \le x) = \ldots = P(Y_j + \Delta \le x)$$

It follows that Y_i + Δ and X_i have the same probability distribution.

Form all m*n pairwise differences X_i - Y_j, and arrange these differences from smallest to largest.
Let pwd(k) be the kth smallest difference, and let k_a and k_b be integers such that k_a < k_b.
Pwd = pairwise differences

The inequality

pwd(k_a) < Δ <= pwd(k_b)

Holds iff at least k_a and no more than k_b-1 of the pairs (X_i,Y_j) satisfy the inequality

X_i - Y_j < Δ

Or equivalently

X_i < Y_j + Δ

Since Y_j+Δ and X_i have the same probability distribution, probabilities involving such inequalities may be obtained from the U statistic.
In particular:

P(at least k_a and no more than k_b-1 of pairs satisfy X_i < Y_j + Δ) = P(k_a <= U <= k_b-1)

Example 2.6.2 the trace element cerium was measured in samples of granite and basalt.

1. For all pairwise differences of the form X_i - Y_j
2. Arrange the pairwise differences in order from smallest to largest
3. X
4. Having chosen k_a and k_b, we form the confidence interval: pwd(k_a) < Δ < pwd(k_b-1)
5. X

If sample sizes allow, we obtain k_a and k_b from Table A4 in the Appendix. For instance if we want a 90% level of confidence.

From Table A4, we find l_.05 = 7, so k_a = 8 and u_.05 = 29, so k_b = 29

Hodges-Lehmann Estimate
The median of all the pairwise differences of the form X_i-Y_j is called the Hodges-Lehmann estimate of Δ.
The Hodges-Lehmann estimate is sometimes suggested as a nonparametric alternative to the difference of the means.