Combining the run and Mann-Whitney-Wilcoxon tests

R. M. Dudley, Feb. 9, 2005, for 18.465

If two tests of the same hypothesis H_0 are done at level α, and just one of the tests rejects H_0, then by a simple Bonferroni correction we could say H_0 is rejected at level 2α. If the two test statistics are independent under H_0, then the precise level is $2\alpha - \alpha^2$, which is close to 2α since α is small.

For the test (called the MWW test) based on the Wilcoxon two-sample rank-sum statistic W_{RS} and the run(s) test based on the number R of runs in the combined sample, R and W_{RS} are not independent. Here W_{RS} is the sum of the ranks of the n Y's in the combined sample and there are m X's. Then R has its smallest possible value 2 if and only if W_{RS} has either its smallest possible value $n(n+1)/2$ or its largest possible value $[(m+n)(m+n+1) - m(m+1)]/2$. For other values, there is dependence although not as strong. Odd values of R tend to make W_{RS} closer to its mean and even values tend to make it farther away as seen especially for $R = 2$.

A combined test will be described which is the run test supplemented by the MWW test, to avoid the Bonferroni correction and keep level α, while also keeping the main advantages of both the runs and MWW tests.

Recall that the run test rejects the hypotheses H_0 that the X_i and Y_j are all i.i.d. with the same continuous distribution $F = G$, for small values of R. Let P_0 denote probabilities and E_0 expectations, assuming H_0 is true. For given m, n, and α with $0 < \alpha < 1$ (for definiteness, $\alpha = 0.05$), let r_0 be the borderline value of R for the runs test at level α, in the sense that $P_0(R < r_0) < \alpha \le P_0(R \le r_0)$. If $R < r_0$, or in the special case that both $R = r_0$ and $\alpha = P_0(R \le r_0)$, we will reject H_0 by the runs test at level α. If H_0 is rejected, we can then apply the two-sided MWW test just to decide whether the data give evidence for a location alternative, with the X's tending to be less than the Y's or vice versa. If the MWW test would not have rejected H_0, and if R is odd, especially if $R = 3$, we can decide that whichever variables (X's or Y's) are in the first and last runs are more dispersed than the others (Y's or X's, respectively).

If $R > r_0$, H_0 is not rejected by the runs or combined test.

The remaining case is where $R = r_0$, the borderline value, and $\alpha < P_0(R \le r_0)$. For example, if $m = n = 6$, we have

$$P_0(R \le 3) = 0.0130 < 0.05 < P_0(R \le 4) = 0.0671.$$

The run statistic R has only 11 possible values in this case, in general $2n - 1$ if $m = n$ or $2\min(m, n)$ otherwise. W_{RS} has $mn+1$ possible values. So R is coarse-grained with large atoms of probability, as just seen with the rather big atom $P_0(R = 4) = 0.052$. We can break such atoms into finer parts and get test levels closer to α using the MWW test.

If $R = r_0$, the combined test being defined here calls for next doing an MWW test. Let w be the observed value of W_{RS}. Let $\mu = E_0 W_{RS} = n(m + n + 1)/2$. The combined test will reject H_0 if $R = r_0$ and

$$P(m, n, w|r_0) \equiv P_0\left(|W_{RS} - \mu| \ge |w - \mu| \big| R = r_0\right) \le (\alpha - P_0(R < r_0))/P(R = r_0).$$

If w is not far enough from μ for the above inequality to hold, then H_0 is not rejected when $R = r_0$. The resulting combined test has level between $P(R < r_0)$ and α and is usually much closer to α than $P(R < r_0)$ is.

The conditional probabilities $P(m, n, w|r_0)$, as functions of four variables, would need to be found by a computer as needed. The unconditional probabilities $P(m, n, w) = P_0(|W_{RS} - \mu| \geq |w - \mu|)$ are available from tables for some w and existing computer packages for general w. The upper bound $P(m, n, w|r_0) \leq P(m, n, w)/P(R = r_0)$ may be helpful: we can reject H_0 if $P(m, n, w) \leq \alpha - P_0(R < r_0)$.

In the example with $m = n = 6$, $\alpha = 0.05$, and $r_0 = 4$, H_0 will be rejected if $R = 4$ and $|w - \mu| \geq 8$, as found by hand calculation. Here $\mu = 39$. The resulting combined test will have level quite close to α. About alternatives, the (unconditional) two-sided MWW test for $m = n = 6$ will reject H_0 at level $\alpha = 0.05$ only if $|w - 39| \geq 13$. If that happens we can decide for a location alternative when $R = 4$. Otherwise, if $R = 4$ and $8 \leq |w - 39| < 13$, we reject H_0 without specifying a type of alternative, because we've done it with the runs and MWW tests combined, not with either one separately.