REPLICATION AND REPEATABILITY - WHEN IS ENOUGH ENOUGH?

Melanie B. Bayles
Laval M. Verhalen
Bruce E. Greenhagen
Bradley R. Barnes
Oklahoma State University
Stillwater, OK

Summary

Researchers trying to detect significant differences among germplasm lines or cultivars for a particular trait are often faced with a dilemma as to the number of replications to use. They must balance the requirement for sufficient replications to determine statistical significance with the very real limitations of space, manpower, time, and funding. This study was designed to provide some insight into the number of replications that should be used in an experiment so it is possible to measure statistical differences for traits with differing degrees of environmental interactions.

Experiments were conducted for 8 years at 1 to 5 locations each year (3 irrigated, 2 dryland). A total of 14 cultivars of upland cotton (*Gossypium hirsutum* L.) were included in the tests with four cultivars used per location. In each case, the experimental design was a RCB with 25 replications. Traits measured included lint yield, selected lint yield components, and HVI fiber properties. This preliminary report includes data from 5 irrigated tests conducted at 3 locations in 2 years. A total of 6 cultivars were included in the 5 tests. Only results for lint yield and fiber strength are presented.

Data from all 25 replications were analyzed using SAS Proc GLM with an alpha level of 0.05. Similar analyses of 100 unique subsets each of 2, 3, 4, 5, 6, 8, 10, 12, 15, and 20 replications were conducted. Since we were interested in looking at the differences between means (i.e., "picking a winner" from among a set of qualitative treatments), it was appropriate to use a multiple comparison test. Fisher's Protected LSD was selected because of its common usage in the literature and in our research program. The 25-replication experiment was considered to be the "standard" for that environment and the subset analyses results were compared to it for conformity of significance vs. nonsignificance and direction of those differences (if significant).

The number of replications required to detect significant differences in lint yield differed greatly and depended largely upon the magnitude of the yield differences between cultivars. For example, no significant differences between cultivars were detected in the 25-replication tests for yield differences of 37 lbs. per acre or less. Using the criterion of 95% correspondence between the 25-replication test vs. those with fewer replications, 20 replications consistently detected differences of 109 lbs. per acre or more. Similarly, 15, 12, and 10 replications did so for 145 lbs. per acre and higher; 8 and 6 replications did so for 170 lbs. per acre and higher; and 5, 4, and 3 did so for 305 lbs. per acre and more. Two replications were inadequate for consistently detecting lint yield differences as large as 457 lbs. per acre.

No significant differences between cultivars were detected in the 25-replication tests for fiber strength values of 0.9 g/tex or less. With 20 and 15 replications, differences of 2.2 g/tex could be consistently detected at the 95% level of correspondence. Twelve replications consistently detected 2.7 g/tex or more; 10 replications, 2.8 g/tex or more; 8 and 6, 3.5 g/tex or more; and 5 and 4, 5.8 g/tex. Three and 2 replications were insufficient to consistently detect fiber strength differences as large as 5.8 g/tex.

Additional data sets (including 8 more irrigated tests and 10 dryland experiments) remain to be examined to determine if the preliminary pattern of results holds true. If it does, larger numbers of replications should probably be used in most experiments. This appears to be true for traits with large environmental effects (such as lint yield) and those that are more stable over environments (i.e., fiber strength). In any case, it appears that two replications are never enough!