INTERPRETATION AND USE OF UNIVERSITY FORAGE VARIETY TRIAL RESULTS

Thomas C. Griggs[1]

ABSTRACT

Forage varieties are tested in university programs to compare production and quality under controlled conditions. For comparisons to be valid, assumptions must be met to remove variation in measurements among plots that is not due to variety differences. Observed differences among varieties are tested to see if they are different from zero at particular levels of significance via analysis of variance. Means of varieties are compared with least significant difference (LSD) or similar tests and coefficient of variation (CV) is used to assess quality control.

Key Words: alfalfa, varieties, yield trials, analysis of variance, LSD, CV

Hay growers, land-grant university extension educators, crop advisors, and seed dealers often consult reports from university forage testing when selecting appropriate forage varieties for environmental and management conditions. Alfalfa variety performance data are available electronically for many states from websites of land-grant universities and the North American Alfalfa Improvement Conference. Performance data are typically used in conjunction with local experience and annually-updated alfalfa fall dormancy and pest resistance ratings published by the Alfalfa Council in paper and electronic formats. The North American Alfalfa Improvement Conference website is also an excellent source of information on genetic background, release dates, and fall dormancy and disease and pest resistance levels of older alfalfa varieties that are not listed in the current Alfalfa Council bulletin.

Forage species and varieties are tested in university programs for the purpose of providing valid comparisons of dry matter production and quality from controlled conditions. Selection of trial locations is based on identification of environmental conditions that are representative of grower conditions, in combination with logistical constraints such as costs of travel and availability of local support. In spite of occasional grower perceptions that unique and unusual substances and management practices are used in university test plots, applications of inputs and management are typically in accordance with current practices of progressive growers. Plot areas are selected and managed to minimize plot-to-plot variation due to environmental conditions and previous field history. Considerable effort is directed at achieving uniformity of fertilizer and irrigation applications. For variety comparisons to be valid, numerous assumptions relating to statistical analysis must be met so that variation in measurements among plots that is not due to variety differences can be factored out. Some of these requirements are replication of the treatments (varieties or 'entries') and randomization of their placement in the plots. An individual entry will commonly be replicated four to six times in an alfalfa trial, e.g., a trial with 25 varieties will have

[1] T.C. Griggs, Extension Agronomy Specialist, Dept. of Plants, Soils, & Biometeorology, Utah State Univ., Logan, UT 84322-4820; Email: tgriggs@ext.usu.edu. **In**: Proceedings, Western Alfalfa and Forage Conference, 11-13 December, 2002, Sparks, NV, UC Cooperative Extension, University of California, Davis 95616 (see http://alfalfa.ucdavis.edu).

100 to 150 plots. While growers often identify more readily with field-scale plots than with typical small plots that may occupy only 36-80 ft^2 each, research plots are sized to meet the constraints of budgets and assumptions underlying experimental designs and statistical analyses.

Plots are harvested with flail and sickle bar mowers to obtain whole-plot fresh weights. Depending on the program, an approximately 200-gram (0.5 lb) subsample is obtained from some or each of the harvested plots for determination of dry matter (DM) concentration in the fresh forage, usually by oven drying at approximately 140° F. In some cases, no subsamples are obtained and a constant DM concentration determined in previous work is assigned to each plot. Plot weights are then scaled to lbs or tons DM/ac and analyzed with statistical software to detect variety differences. The following table presents example data from an irrigated trial in northern Utah. To keep the table brief, only the highest- and lowest-producing six of the original entries are shown. As shown in the last column, annual updates rank the entries on the basis of DM production averaged or totaled over the production years of the trial. Rankings of entries within any one year are not necessarily the same as over the length of the trial, pointing to the value of data from multiple years in predicting the long-term performance of a variety under conditions similar to those of the trial.

Utah State Univ. Kaysville irrigated alfalfa 1999-2001 (partial data for example only).
Entries ranked by 3-yr mean dry matter (DM) production.
Seeded 9/16/98 @ 20 lb bulk seed/ac, 5 rows @ 6-in spacing, Kidman fine sandy loam.
29 entries in randomized complete block design with 5 replicates, plot size 3 x 20 ft.

	Total	Total	2001 DM production by harvest date					Total	Seas. tot.,
Entry (cultivar)	1999	2000	May 17-18	June 27	July 30	Sept. 1	Oct. 20	2001	3-yr mean
			---- Tons DM/ac ----						
ZX 9451	8.08	11.88	2.92	2.93	1.89	1.78	1.22	10.74	10.23
DK 142	8.26	11.43	3.00	2.83	2.09	1.55	1.11	10.58	10.09
Magnum V	8.11	11.65	2.83	2.78	2.12	1.44	1.10	10.26	10.01
Deseret	7.79	11.65	3.12	2.83	2.14	1.42	0.93	10.47	9.97
Archer II	7.97	11.02	2.97	2.82	2.02	1.80	1.31	10.92	9.97
54Q53	8.20	11.34	3.06	2.80	1.84	1.49	1.05	10.24	9.93
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
Reno	6.28	11.33	2.95	2.55	1.69	1.55	1.00	9.73	9.11
Emperor	6.93	10.54	2.80	2.55	1.78	1.36	0.97	9.46	8.98
Ripin	6.50	10.06	2.75	2.67	1.90	1.35	0.89	9.56	8.71
TMF Multiplier II	6.35	9.97	2.73	2.41	1.92	1.47	1.04	9.58	8.63
Rambo	6.35	10.03	2.63	2.50	1.82	1.27	0.86	9.08	8.49
Ranger	5.88	9.57	2.65	2.30	1.53	1.31	0.84	8.63	8.03
Mean	7.23	10.87	2.87	2.66	1.90	1.48	1.03	9.94	9.35
LSD (0.05)	NS*	0.86	0.27	0.19	0.25	0.26	0.17	0.64	0.60
LSD (0.30)	NS	0.45	0.14	0.1	0.13	0.13	0.09	0.34	0.32
CV (%)	17.2	6.5	8.2	6.2	11.5	15.4	14.4	5.6	8.9

*NS: no significant differences (0.05) among entries.

Although small-plot data are usually presented on an oven-dry, rather than air-dry, basis, DM production is perhaps 10-20 percent higher than would be obtained by commercial harvesting and baling of the same material, because the usual DM losses due to respiration in the windrow, machinery traffic, and leaf shatter from mechanical handling are not encountered. Apparent differences among varieties within a cutting, year, or life of the trial are tested to see if they are different from zero at a particular significance level via analysis of variance. Testing for treatment differences at a commonly-used significance level of 5 percent simply means that for entries that are truly not different, apparent differences as large as those observed would occur by chance 5 percent of the time. Apparent differences among entries that do not differ statistically at a level of 5 percent, or whatever higher level the researcher may be comfortable with, are those that would occur relatively frequently by chance alone, e.g., more than 5 percent of the time. Apparent differences that are not significant are due to some combination of random variation or inadequate experimental design or technique, and are indicated by 'NS' in the row of LSD values in the table. The statistical interpretation of 'NS' is that at a given level of probability, variations among entries are not significantly different from zero, i.e., the varieties perform the same or there is insufficient statistical power to detect differences. Lack of differences is more likely to occur in first production years or under conditions of environmental stress or pest pressure.

If significant differences among varieties were detected by statistical analysis, the least significant difference (LSD) or other mean-separation procedure is used to indicate which adjacent means are statistically different from each other. The LSD is interpreted as the minimum numerical difference required between adjacent means of entries in a column in order to declare them different at a given level of significance. The table shows LSD values in units of tons DM/ac, for 5 and 30 percent probability levels of incorrectly declaring differences among adjacent means when true differences do not exist. Different significance levels are provided to allow users a choice of more conservative or liberal interpretations of statistical tests.

A final statistic presented in the table is the coefficient of variation (CV), which expresses experimental error as a percentage of the mean performance of all plots. Experimental error represents variation among plots that is not related to variety differences, but instead to inherent variability among plots treated similarly and to inconsistency and mistakes in experimental protocol and measurements. The CV is interpreted as an index of the degree of precision available for detecting differences among entries, and values below 6-10 percent are often used as evidence of good quality control in well-conducted trials.

References:

Fall Dormancy & Pest Resistance Ratings for Alfalfa Varieties, 2002/2003 ed., Alfalfa Council, Kansas City, MO, http://www.alfalfa.org/.

Internet links to alfalfa variety trial yield results. North American Alfalfa Improvement Conf., http://genes.alfalfa.ksu.edu/stdtests/.

Sheaffer, C., et al. 2002. Forage yield testing guidelines. 3 pp. *In* C.C. Fox et al., ed. Standard tests to characterize alfalfa cultivars, 3rd ed. North American Alfalfa Improvement Conf., http://genes.alfalfa.ksu.edu/stdtests/.