**Relationship between Mean Square Errors and Wheat Grain Yields in Long-term Experiments**

Abstract

When yield data from field experiments are combined, the impact of environment on treatment is ignored. The objectives of this work were to determine the frequency for combining year to year data. Grain yield mean square errors (MSE) were calculated by-year from two long-term winter wheat experiments (E222 and E502). These trials have been continuously planted to winter wheat (Triticum aestivum L), and both annually receive different nitrogen (N), phosphorus (P) and potassium (K) fertilizer rates. Relationships between average grain yields, coefficients of variation (CV) and MSE were evaluated over 42 and 41 years for E222 and E502, respectively. The F-statistics (large MSE/small MSE, 39dfn, 39dfd, alpha=0.10) computed from two and three consecutive years were consistently more than the minimum F-value required to combine sites. For 83 site years of data, combining any two or three consecutive year-periods was not advisable. Individual years should be analyzed and reported on independently.

Shortened version of the title: Mean square errors and wheat grain yields

Key words: mean square errors, wheat, grain yield, environment, heterogeneity of error variance,

Abbreviations:
AOV, analysis of variance
CIMMYT, International Maize and Wheat Improvement Center
CV, coefficient of variation, %
LSD, least significant difference
MSE, mean square error

Introduction

In long-term experiments, where the same treatments are applied to specific plots, year after year, interpretation of yield response is often conducted over longer periods of time without considering year-to-year homogeneity/heterogeneity of experimental error.

Since the advent of modern statistics, experimental errors have long been considered and discussed. Early work by Mercer and Hall (1911) noted that the degree of confidence which may be attached to the results from field experiments depends on experimental error. Mercer and Hall (1911) further reported that experimental error could be reduced by repeating the experiment over a long period of time. This coming from one of the oldest long-term experiments in the world at Rothamsted Research, UK ([www.rothamseted.ac.uk](http://www.rothamseted.ac.uk), visited March 22, 2015).

Fisher (1926) reported that estimates of field errors coming from any specific experiment may or may not be a valid estimate of the actual errors affecting the averages, or differences of averages, and that are required to estimate experimental error. Fisher (1926) goes on to state that the estimate of error afforded by the “replicated trial” depends on differences between plots treated alike. The relevance of this early work was further reflected in his noting that a valid estimate of error will be obtained by arranging the plots in a deliberately random fashion. The value of replication (provides the opportunity to estimate experimental error) and randomization (provides the opportunity to have an unbiased estimate of experimental error) were thus cemented into the chronicles of biometry.

Recent comments from Dr. Jose Crossa noted that “analyzing data over locations and/or years can be problematic as this can ignore treatment by environment interactions that are prevalent in agricultural studies” (personal communication March 23, 2015, Dr. Jose Crossa, Head Statistics, CIMMYT, Mexico, DF). Because most field experiments are conducted over two or more years/locations, McIntosh (1983) formulated recommendations for fixed, mixed, and random models.

The objectives of this study were to determine the relationship between MSE and yield level, the year to year variability in MSE’s and the frequency for combining year to year data from long-term experiments.

Materials and Methods

Two long-term winter wheat experiments, Experiment 222 (Stillwater, OK) and Experiment 502 (Lahoma, OK) were used to determine the efficacy/feasibility of combining year to year grain yield data from long-term experiments. Experiments 222 and 502 were established in 1969 and 1971, respectively, under conventional tillage. Both experiments were converted to no-tillage management in 2011. No yield data was collected in 1969, 1975, and 2009 for Experiment 222 and 1973 and 2007 for Experiment 502, all due to drought. Thus, for this analysis, 42 and 41 years of data were included for Experiments 222 and 502, respectively.

Experiment 222 is located at the Agronomy Research Station in Stillwater, OK at an altitude of 272 masl, on a well-drained, deep and slowly permeable Kirkland silt loam (fine, mixed, thermic Udertic Paleustolls). Experiment 502 is located at the North Central Research Station in Lahoma, OK at an altitude of 396 masl on a well-drained, deep and moderately permeable Grant silt loam (fine-silty, mixed, thermic Udic Argiustolls) soil.

The average annual precipitation at Stillwater and Lahoma is 80 cm. The experimental design for both trials is a randomized complete block with four replications. There were 12 and 13 treatments that comprised application of different rates of N, P, and K fertilizers in both Experiment 222 and Experiment 502. This resulted in 36 and 39 degrees of freedom in the error term when each was analyzed by year. Nitrogen, P, and K were applied as urea (45% N), triple super phosphate (20% P), and potassium chloride (53% K), respectively, in both experiments. Plots were permanent and received fixed rates of N, P, and K, every year. Individual plots at Stillwater are 6.1 m wide and 18.3 long, and at Lahoma are 4.9 m wide and 18.3 m long. Experimental plots were conventionally tilled every year in the summer to a 15-cm depth using a disk plow. From establishment to 2010, plots were tilled before fertilization using a spike tooth harrow. However, in the fall of 2011, both experiments were converted to no-till. Winter wheat was planted in 25.4-cm wide rows at seeding rates of 67 kg ha-1. In some years, the seeding rate was increased to 110 kg ha-1 in anticipation of poor germination and emergence due to unfavorable soil moisture conditions at seeding. Since 1992, winter wheat has been planted in 19.1-cm rows at Stillwater. At both sites, varieties were changed with time due to increased genetic yield potential and the need for rust resistance (Table 1). In both experiments, preplant fertilizer was broadcast and incorporated in late August to mid-September. Since 2011, fertilizers have not been incorporated. Winter wheat was planted in late September to early October every year. Grain yield data was collected in each plot and year using a self-propelled combine.

For both long-term data sets, analysis of variance was conducted by-year, accounting for replication and treatment in the model. Mean yields, and mean square errors (MSE) were recorded, and coefficients of variation (CV) determined accordingly. Linear relationships between mean grain yields, MSE, CV, year, and computed Least Significance Difference (LSD, alpha = 0.05) tests were also evaluated. Bartlett’s (Bartlett, 1937) test for evaluating homogeneity of MSE terms from the by-year AOV’s was also performed, both year to year, and over three year sequences (F = large MSE/small MSE, df numerator, df denominator). For this, df represents the degrees of freedom in the error term from the by-year analysis. For Experiments 222 (36 df) and 502 (39 df), F statistics at an alpha level of 0.05 were 1.70 and 1.74, respectively. The null hypothesis was that there were no differences in the by-year error terms. F statistics were plotted as a function of year. Comprehensive discussion of both Experiment 502 and Experiment 222 can be found in the following publications (Westerman et al., 1994, Raun et al., 1998, and Raun et al., 2011).

Results

Grain yield and MSE

For both Experiment 222 and Experiment 502, MSE’s tended to increase as yields increased (Figures 1 and 2). Numerically this makes sense as the actual numbers on both sides (yield and MSE) would be expected to be larger. Nonetheless, these relationships were weak (coefficient of determination, r2, both less than 0.19).

MSE and Year

Mean square errors from by-year analysis of variance on wheat grain yields showed no measurable change in Experiment 222, but increased slightly in Experiment 502 with time (Figures 3 and 4).

Year and CV

At both locations, the coefficient of variation (CV = (square root (MSE/experimental mean) \*100)) from the analysis of variance model showed no change with time (Figures 5 and 6). This would suggest that the experimental errors associated with conducting these trials remained relatively constant, despite known influences of the environment on treatment response. This was encouraging because it suggests that the methods (planting, tillage, harvest, weed control, other) employed when conducting these trials were repeatable and also because CV’s averaged less than 12 and 16 for experiments 222 and 502, respectively, and this over a notable number of years. At both locations, CV’s decreased with increasing yield level, and that was consistent with findings from Taylor et al. (1999).

Combining Ensuing 2 and 3 Year Data

F-tests computed (large MSE/small MSE) for two and three ensuing year periods, in Experiment 222 and Experiment 502, are reported in Figures 7-10, respectively.

For Experiment 222 consecutive 2-year data should not have been combined 23 of 42 possible combinations or 55% of the time, and this value rose to 90% when 3-year periods were considered (Figures 7 and 8, F value of 1.74 needed to declare significantly different).

This same analysis for Experiment 502, showed that 18 of 41 possible combinations of 2-year consecutive year periods had F values in excess of 1.7. In other words, ensuing two year periods should not have been combined 43% of the time (Figure 9). For consecutive 3-year periods (1971-2014), F values exceeded 1.7, 73% of the time, further restricting combining data over years (Figure 10).

Conclusions

Comprehensive data from two long-term experiments analyzed show that combining any two, or three year consecutive data sets is not advisable based on the failure to meet homogeneity of error variance tests. This also suggests that combining year to year data can overlook the importance that environment has on treatment response. Furthermore, computed F values for 2 and 3 year consecutive periods document that the influence of environment from one year to the next was never the same.

References

Bartlett, M.S. 1937. Properties of sufficiency and statistical tests. Proc. Royal Society of London. Series A, Mathematical and Physical Sciences. 160:268-282.

Fisher, R.A. 1926. The arrangement of field experiments. Journal of the Ministry of Agriculture. 33:503-514.

McIntosh, M.S. 1983. Analysis of combined experiments. Agronomy Journal. 75:153-155.

Mercer, W.B., and A.D. Hall. 1911. The experimental error of field trials. Journal of Agricultural Science. 4:107-132.

Raun, W.R., G.V. Johnson, S.B. Phillips and R.L. Westerman. 1998. Effect of long-term nitrogen fertilization on soil organic C and total N in continuous wheat under conventional tillage in Oklahoma. Soil & Tillage Research. 47:323-330.

Raun, William R., John B. Solie, and Marvin L. Stone. 2011. Independence of yield potential and crop nitrogen response. Precision Agriculture. 12:508-518.

Taylor, S.L., M.E. Payton and W.R. Raun. 1999. Relationship between mean yield, coefficient of variation, mean square error and plot size in wheat field experiments. Communications Soil Science and Plant Analysis. 30:1439-1447

Westerman, R.L, R.K. Boman, W.R. Raun and G.V. Johnson. 1994. Ammonium and nitrate nitrogen in soil profiles of long-term winter wheat fertilization experiments. Agronomy Journal. 86:94-99.

Figure 1. Relationship between mean square errors (MSE) from by-year analysis of variance and mean grain yield level, Experiment 222, 1969 to 2014.

Figure 2. Relationship between mean square errors (MSE) from by-year analysis of variance and mean grain yield level, Experiment 502, 1971 to 2014.

Figure 3. Relationship between mean square errors (MSE) from by-year analysis of variance and year, Experiment 222, 1969 to 2014.

Figure 4. Relationship between mean square errors (MSE) from by-year analysis of variance and year, Experiment 502, 1971 to 2014.

Figure 5. Relationship between the coefficient of variation (CV) from by-year analysis of variance and year, Experiment 222, 1969-2014.

Figure 6. Relationship between the coefficient of variation (CV) from by-year analysis of variance and year, Experiment 502, 1971-2014.

Figure 7. F-statistic (large MSE/small MSE) plotted over time for 2 ensuing years, Experiment 222, 1971-2014. F-statistics above the red line (36 dfn, 36 dfd, 0.05 = 1.74) note 2 year periods where data should not have been combined.

Figure 8. F-statistic (large MSE/small MSE) plotted over time for 3 ensuing years, Experiment 222, 1971-2014. F-statistics above the red line (36 dfn, 36 dfd, 0.05 = 1.74) note 3 year periods where data should not have been combined.

Figure 9. F-statistic (large MSE/small MSE) plotted over time for 2 ensuing years, Experiment 502, 1971-2014. F-statistics above the red line (39 dfn, 39 dfd, 0.05 = 1.70) note 2 year periods where data should not have been combined.

Figure 10. F-statistic (large MSE/small MSE) plotted over time for 3 ensuing years, Experiment 502, 1971-2014. F-statistics above the red line (39 dfn, 39 dfd, 0.05 = 1.70) note 3 year periods where data should not have been combined.

Table 1. Winter wheat varieties planted in Experiments 502 and 222, 1969-present.

|  |  |  |
| --- | --- | --- |
| Year | Experiment 502 | Experiment 222 |
| 1969 | - | Scout66 |
| 1970 | - | Scout66 |
| 1971 | Nicoma | Scout66 |
| 1972 | Nicoma | Scout66 |
| 1973 | Nicoma | Scout66 |
| 1974 | Nicoma | Scout66 |
| 1975 | Triumph64 | Triumph64 |
| 1976 | Triumph64 | Triumph64 |
| 1977 | Osage | Triumph64 |
| 1978 | Triumph64 | Osage |
| 1979 | TAM101 | Osage |
| 1980 | TAM101 | Osage |
| 1981 | TAM101 | TAM101 |
| 1982 | TAM101 | TAM101 |
| 1983 | TAM101 | TAM101 |
| 1984 | TAM101 | TAM101 |
| 1985 | TAM101 | TAM101 |
| 1986 | TAM101 | TAM101 |
| 1987 | TAM101 | TAM101 |
| 1988 | TAM101 | TAM101 |
| 1989 | TAM101 | TAM101 |
| 1990 | TAM101 | TAM101 |
| 1991 | TAM101 | TAM101 |
| 1992 | TAM101 | TAM101 |
| 1993 | Karl | Karl |
| 1994 | Karl | Karl |
| 1995 | Tonkawa | Tonkawa |
| 1996 | Tonkawa | Tonkawa |
| 1997 | Tonkawa | Tonkawa |
| 1998 | Tonkawa | Tonkawa |
| 1999 | Tonkawa | Tonkawa |
| 2000 | Custer | Custer |
| 2001 | Custer | Custer |
| 2002 | Custer | Custer |
| 2003 | Custer | Custer |
| 2004 | Custer | Custer |
| 2005 | Overley | P2174 |
| 2006 | Overley | Endurance |
| 2007 | Overley | Endurance |
| 2008 | Overley | OKField |
| 2009 | Endurance | OKField |
| 2010 | Bullet | GoLead |
| 2011 | Bullet | Centerfield |
| 2012 | Billings | Endurance |
| 2013 | Billings | OK9935C |
| 2014 | Ruby Lee | Doublestop-CL |