

Kansas State University Libraries

New Prairie Press

---

Conference on Applied Statistics in Agriculture

1989 - 1st Annual Conference Proceedings

---

## A COMBINED ANALYSIS OF EXPERIMENTS WHEN TREATMENTS DIFFER AMONG EXPERIMENTS

Paul N. Hinz

Mario R. Pareja

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

---

### Recommended Citation

Hinz, Paul N. and Pareja, Mario R. (1989). "A COMBINED ANALYSIS OF EXPERIMENTS WHEN TREATMENTS DIFFER AMONG EXPERIMENTS," *Conference on Applied Statistics in Agriculture*.  
<https://doi.org/10.4148/2475-7772.1465>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact [cads@k-state.edu](mailto:cads@k-state.edu).

## A COMBINED ANALYSIS OF EXPERIMENTS WHEN TREATMENTS DIFFER AMONG EXPERIMENTS

Paul N. Hinz and Mario R. Pareja  
Department of Statistics  
Iowa State University, Ames, IA  
and CATIE, Turrialba, Costa Rica

### Abstract

The advantages of repeating experiments in several locations and years are discussed and standard methods of analysis are reviewed. The methods assume that the same treatments are used in each experiment. This paper discusses a method used for a combined analysis when the treatments represent levels of a quantitative factor but differ among experiments. The method makes use of multiple regression analysis in which a continuous variable represents treatment levels, classification variables represent experiments, and products of the continuous and classification variables represent differences among experiments. The method is illustrated on data from a series of experiments designed to study the relationship of grain yield of soybeans as affected by the density of the weed species velvetleaf. The analysis determined that yield loss was linearly related to weed density but that the slope of the relationship differed among years. The slope differences were correlated with August rainfall, and a model is suggested that accounts for both within-experiment variability due to weed density and between-experiment variability due to August rainfall.

Keywords: combined experiments, multiple regression, velvetleaf,  
soybean

### 1. Introduction

In agricultural research, experiments are often repeated in different locations and years. Indeed, the practice is almost required if you wish to publish in certain journals. One objective of repeating an experiment is to obtain treatment means that will represent the average response of treatments under varying conditions. This is important if a treatment is to be adopted for general use. A second objective is to discover how conditions external to the experiment affect the treatments. To achieve this objective, measurements must be made of variables that describe the conditions under which each experiment is performed and the changes in treatment effects related to the variables. The types of conditions that might affect outcomes of agricultural experiments are soil type, insects, disease, and weather. This objective can be difficult to achieve because the variables that explain experiment differences are often unknown or can't be measured.

The analysis of a series of experiments is described in Cochran and Cox (1957). Treatment means of each experiment are regarded as responses from a randomized block experiment in which the experiments are equated to blocks.

The mean square for treatment by experiment interaction is used as an estimate of experimental error to make inferences for treatments. A rationale for using this mean square is that the estimate of experimental error for treatment means which came from a combined analysis should include a component that accounts for variability among experiments. The mean square for the treatment by experiment interaction contains such a component and is therefore useful for accounting for experiment differences due to factors such as soil type, insect damage, disease, and weather. Experimental error within experiments can be averaged or pooled to obtain what is called a "pooled error." This pooled error is used to make inferences about the treatment by experiment interaction.

This is a bare-bones description of a combined analysis. This analysis would be enhanced by use of single-degree-of-freedom contrasts and other such data-analysis techniques. Also, complications arise. Cochran and Cox give advice on what to do if variances are nonhomogeneous or if the number of replications or experimental designs differ among experiments. However, it is assumed that there is a common core of treatments used in each experiment.

Although it would seem to make no sense to combine experiments when treatments differ among experiments, an exception would be when all treatments represent different levels of the same quantitative factor. Examples would be amount of herbicide, width of planting, or concentration of an insecticide. The purpose of this paper is to describe the analysis of a series of four experiments for which the treatments differed among experiments but were all levels of a single quantitative factor. These experiments concern the yield of soybeans and how the yield is affected when weeds are present. The treatments are weed densities, which can be described quantitatively as number of plants/ha. The basic analysis will be described. The analysis will be extended to describe differences among experiments and how those differences could be explained by external conditions.

## 2. Combined Analysis for Quantitative Treatments

The type of analysis used when data come from a series of experiments can be illustrated by supposing that data are collected from three experiments where the same number of replications of the same four treatments are used in each experiment. The analysis can be performed using the treatment means from each experiment. Experiments and treatments can be regarded as cross-classified factors because the treatments are assumed to be the same in each experiment. This means that the 12 treatment means from the illustration can be analyzed in an analysis of variance of the form shown in Table 1. Alternatively, single observations could be used in the analysis rather than treatment means. If each experiment used a randomized complete block design, a complete analysis of variance table for single observations would include the sources of variation shown in Table 1 as well as sources due to blocks nested within experiments and pooled error. The analysis of variance based on treatment means is presented because it is less complicated than the analysis of single observations. In addition, no essential information is lost when treatment means are analyzed because the essential parts of the analysis of treatment means can be easily converted to a single observation basis merely by multiplying the sums of squares by  $\frac{1}{r}$ , the number of replications of each treatment in an experiment.

An informative subdivision of the sums of squares in Table 1 is possible if the treatments represent levels of a quantitative factor such as amount of fertilizer or concentration of herbicide. When the same treatment levels are used in each experiment, the sums of squares for the main effect of treatment and experiment by treatment interaction can be subdivided to account for possible linear relationships of the response with the level of the treatment factor. The form of this analysis of variance table is shown in Table 2. If linear relationships do not fit the data well, an adequate analysis might be obtained by including subdivisions for quadratic and experiment by quadratic components in the analysis of variance table. Other response functions could be used in cases where quadratic functions do not adequately describe the data.

When linear response functions describe the data adequately, the sources of variation in Table 2 can be rearranged to emphasize the linear regression as shown in Table 3. The interpretation of the analysis shown in Table 3 can be explained by assuming that a separate regression equation is obtained for each of the three experiments where the response is regressed onto the factor level of the treatment. The differences among the three intercepts and among the three slopes are analyzed in Table 3. The main effect of experiments provides information about intercept differences and the experiment by linear interaction measures differences among slopes. The linear source can be used to test if the average slope differs from zero.

The benefit of presenting the analysis as shown in Table 3 is that the sources of variation correspond to independent variables in a multiple regression in which experiments are classification variables and treatments are a continuous variable which uses the level of the factor as its value. This means that it is not necessary to express treatments as a classification variable. Further, because treatments can be described as a continuous variable, the multiple-regression approach can be used to analyze data even if the treatment levels differ among experiments. This is the approach used in the following example.

### **3. The Example**

A randomized block experiment was conducted in each of the years 1975, 1976, 1979, and 1980, and the treatments were density levels of the weed, velvetleaf, in soybeans. For the 1975 experiment, the treatments were 0, 3175, 4233, and 7407 velvetleaf plants/ha. Each treatment was replicated once in each block except for the control, or 0 plants/ha., which was replicated twice. The other three experiments were conducted similarly except that the weed densities and number of blocks varied among experiments. A summary of the experiments is shown in Table 4.

The treatment means from these experiments were analyzed by use of the multiple-regression technique described in the preceding section. Use of treatment means in the analysis avoids the complications due to unequal treatment replication both within and between experiments and corresponds to the use of the analysis of unweighted means described in Chapter 14 of Cochran and Cox. The analysis determined that there are significant differences among intercepts and slopes. It is easy to recognize these differences in the plot of the treatment means shown in Figure 1.

#### 4. Extension of Analysis

The goal of the research and data analysis was to provide a way to predict yield reduction due to competition from velvetleaf. The analysis was successful in that a linear regression was found to describe the relationship between yield and weed density but unsuccessful in providing a way to predict yield loss because the slope of the regression differs among years. For this reason, the data were re-expressed to see if yield loss could be predicted in terms of percent yield reduction. For any given year, percent yield reduction was calculated according to the formula

$$\text{REDUCT} = \frac{\text{CONTROL YIELD} - \text{TREATMENT YIELD}}{\text{CONTROL YIELD}} \times 100.$$

The intercept of the linear regression of yield versus weed density was used as the value of the control yield in this formula wherein the intercept was calculated separately for each experiment.

The values of percent yield reduction were analyzed by use of the same multiple-regression method used to analyze the treatment means. The results of the analysis were that slopes differed significantly among years and that no evidence of nonlinear relationships was detected. Unfortunately, for some of the experiments, the intercepts were greater than zero. This implies that these regressions predict reductions greater than zero when weed density equals zero. Also, slopes that differ among experiments means that percent reduction cannot be predicted satisfactorily. A plot of percent reduction versus weed density is shown in Figure 2.

#### 5. Modeling Year Differences

To improve the predictive ability of the multiple regression analysis, it is necessary to explain the slope differences among years. It was noted that 1976 was an exceptionally dry year and, correspondingly, that the slope of the percent reduction versus weed density regression was relatively large. On the other hand, 1979 and 1980 were years of adequate to excess moisture, and the corresponding slopes were relatively small. After a number of different ways of measuring plant available moisture were investigated, it was found that August rainfall was correlated best with the slope of the percent reduction versus weed density regression. A plot of slope versus August rainfall is shown in Figure 3.

August rainfall can be used to provide a very reasonable explanation of yearly differences. August rainfall represents available moisture during the critical period when the soybean pods are filling. If moisture is limited, the beans and weeds are competing for a limited resource, and hence, the weeds deprive the beans of needed moisture. If moisture is plentiful, the competition between beans and weeds is reduced and the yield is not as greatly affected. However logical and appealing this explanation is, we should be aware that it is based on data from four years only. Also, the measure of moisture, August rainfall, was chosen by sorting through a number

of different variables. So, if the explanation of yearly differences is to be taken seriously, it should be confirmed with more experiments of the same type and possibly different kinds of experiments as well.

The preceding analyses suggest that August rainfall (RAIN) and weed density (DENSITY) can be used to explain percent yield reduction (REDUCT) by use of the general model

$$\text{REDUCT} = b_0 + b_1 * \text{DENSITY} + b_2 * \text{RAIN} * \text{DENSITY} + e$$

where  $b_0$ ,  $b_1$ , and  $b_2$  represent unknown parameters to be estimated and  $e$  represents error. Unfortunately, when  $b_0$  is different from zero, this model has the undesirable property of predicting a percent reduction of  $b_0$  when weed density is zero. To overcome this objection, a model where the parameter  $b_0$  was deleted and where the variable DENSITY was replaced with  $\text{DENSITY}^a$  was used to fit the data. The variable  $\text{DENSITY}^a$  with  $a=1$  was not used in the model because it resulted in a significant lack of fit. The use of the power  $a = 3/4$  gave a model that fit the data adequately. The power  $a = 3/4$  was selected by fitting the model

$$\text{REDUCT} = b_1 * (\text{DENSITY})^a + b_2 * \text{RAIN} * (\text{DENSITY})^a + e$$

for values of  $a$  equal to 0 (or log),  $1/3$ ,  $1/2$ ,  $3/4$ , and 1 and then choosing the value of  $a$  that minimized the residual mean square.

## **6. Concluding Remarks**

The analysis of data from these experiments resulted in a model that describes the relationship of percent yield reduction with the variables weed density and August rainfall. Although the model successfully describes the data from the experiments, its use as a prediction equation has some limitations. First, the model includes a variable, August rainfall, that can be determined with certainty only after most growth has stopped and, hence, after yield has already been determined. This means that weed control decisions based on the model are limited by the ability to predict August rainfall. Second, the prediction equation is based on only four years of data, and therefore, the effective sample size for estimating the regression coefficient  $b_2$  is only four. Third, standard errors for the regression coefficients of this type of model can be calculated by use of the methods described by Pantula et al. (1985). However, the methods are not valid for this study because the variable, August rainfall, was selected through a search procedure, and the methods do not account for variables selected in this way.

The importance of repeating experiments should be obvious from the results of the data analysis. The relationship of yield loss with plant-available moisture would not have been discovered had data from only one experiment been available. Also, a reliable assessment of the relationship between yield loss and plant available moisture clearly requires data from a number of different years.

## 7. Summary

An analysis of data from a series of experiments where treatments differ among experiments is discussed. The analysis makes use of a multiple regression model and represents a minor modification of the standard analysis of variance used when treatments are the same in all experiments. The analysis is illustrated on data from a series of four experiments designed to study the relationship of soybean grain yield to velvet leaf density. The variable August rainfall was found to account for differences between experiments and the multiple regression model was extended to account for this variable.

**Table 1. Outline of the analysis of variance of means of four treatments from three experiments.**

<u>Source</u>	<u>d.f.</u>
Experiment	2
Treatment	3
<u>Experiment * Treatment</u>	<u>6</u>

**Table 2. Outline of the analysis of variance of means of four treatments from three experiments showing subdivisions of degrees of freedom when treatments are quantitative.**

<u>Source</u>	<u>d.f.</u>
Experiment	2
Treatment	3
Linear on Level	(1)
Remainder	(2)
Experiment*Treatment	6
Experiment*Linear	(2)
<u>    Experiment*Remainder</u>	<u>(4)</u>

**Table 3. Outline of the analysis of variance of means of four treatments from three experiments showing subdivisions for intercepts and slopes of linear regressions.**

<u>Source</u>	<u>d.f.</u>
Experiment or Intercept Differences	2
Linear or Common Slope	1
Experiment*Linear or Slope Differences	2
<u>Residual or Error</u>	<u>6</u>

Table 4. Experimental designs used for a series of experiments to study how soybean yield is affected by velvetleaf density.

Year	Blocks	Treatments (plants/ha. of velvetleaf)
1975	6	0, 3175, 4233, 7407
1976	5	0, 1112, 2224, 4448, 8896
1979	4	0, 1112, 1481, 2224, 4448, 8896
1980	5	0, 1190, 1587, 2381, 4762, 9524

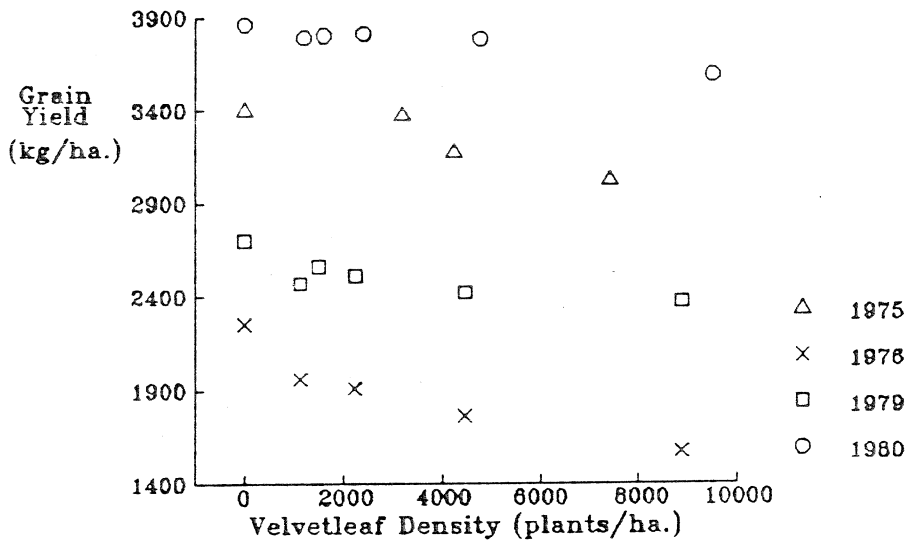


Figure 1. Treatment means for soybean yield versus velvetleaf density for experiments in 4 different years.

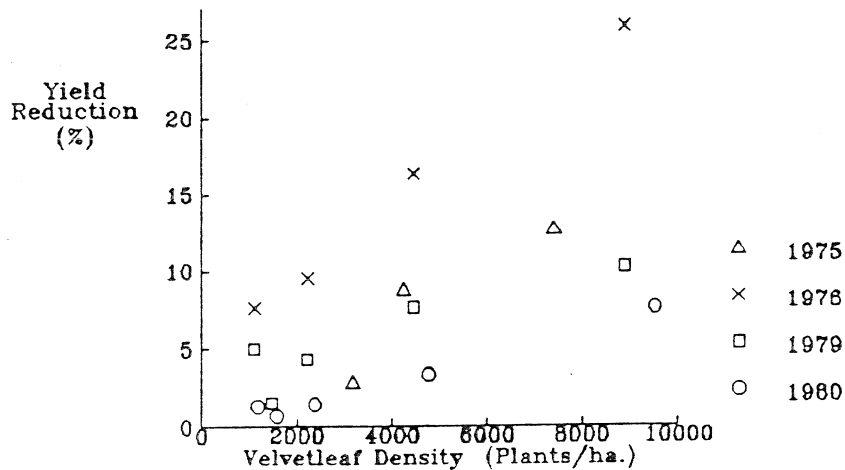


Figure 2. Percent yield reduction of soybeans versus velvetleaf density for experiments in 4 different years.



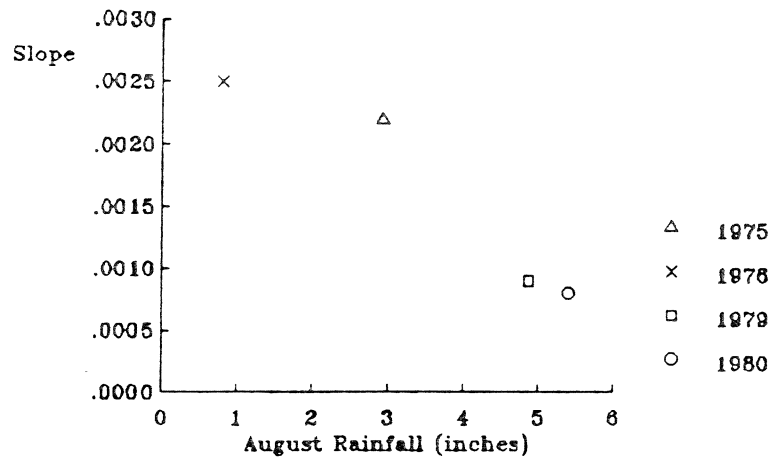


Figure 3. Slope for the linear regression of percent yield reduction versus velvetleaf density plotted against August rainfall.

### References

Cochran, W. G. and G. M. Cox (1957). Experimental Designs. John Wiley and Sons, New York, New York.

Pantula, S. G., L. A. Nelson, and R. L. Anderson (1985). Estimation of linear models for field experiments. Communications in Statistics, Part A-Theory and Methods, 14(9), 2199-2217.

Journal Paper No. J-13574 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa, Project No. 0101.